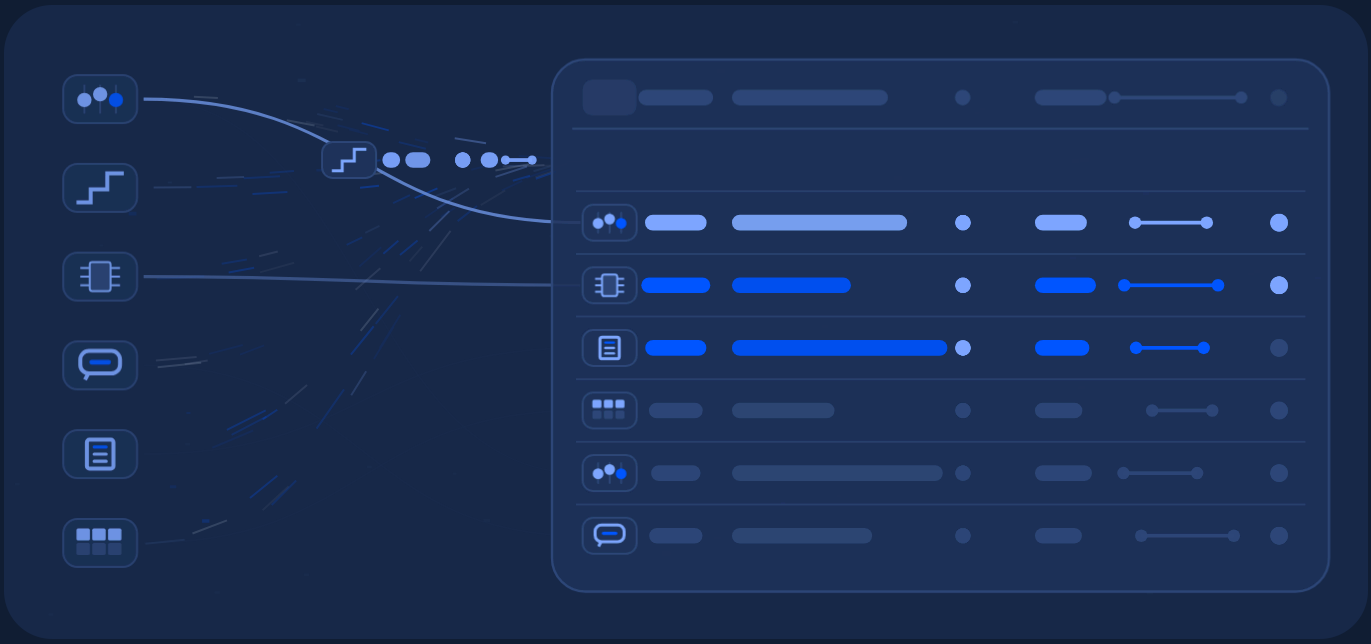
Rillor

EXPLAINER

Due diligence questions to ask a data vendor

Data vendor due diligence, question by question: who the vendor is, where its data comes from, how it checks its suppliers and how to test its answers.



Rillor

Contents

01 Summary

02 Background

03 Method

04 Findings

05 Implications

06 Limits of this analysis

07 Sources

Summary

Data vendor due diligence tests the company behind a dataset, and enforcement orders, regulators' guidance and industry questionnaires supply its questions.

- A vendor's questionnaire answers are claims to test. In 2021 the SEC found that an alternative data provider and its founder made false and misleading representations, some "in response to diligence questionnaires", and both settled without admitting or denying the findings [1, 2].
- Data can pass through several vendors before it reaches a buyer. In a 2014 FTC study, the nine data brokers studied obtained most of their data from other data brokers, and seven of the nine provided data to each other [8].
- Two FTC orders set out a supplier check: a written program, an assessment within 30 days of a new supplier and then every year, records of suppliers' responses, and no use of data collected without consent [5, 7].
- Who the vendor is takes several checks: whether it originates the data or pulls it in from other sources [15], the legal entity and its trade names in California's data broker registry [21, 22], owners and sanctions [28], and five years of suits, subpoenas and regulator inquiries [12].
- For personal data, ask for evidence of consent. France's data protection authority wrote in 2024 that the obligation to prove consent "cannot be fulfilled by the mere presence of a contractual clause" [25].
- Answers need checks the buyer runs itself and a written record. A buyer's guide from FISD, a financial information industry association, lists reviewing litigation history and news and keeping a record of diligence steps taken [16], and SEC examination staff observed advisers that could not show their policies had been consistently implemented [3].
- Voluntary due diligence questionnaires from FISD, a financial information industry association, added questions in 2024 on how a vendor uses generative tools to produce its data [11, 12, 13].
- Diligence continues after signing: annual supplier reassessment in two FTC orders [5, 7], at least 90 days' notice of major changes in the service level recommendations of FISD, a financial information industry association [19], and supplier monitoring over the whole relationship in NIST's Cybersecurity Framework 2.0 [30].

Background

Why buyers review the vendor as well as the data

Data often reaches a buyer through several companies. In its 2014 study of nine data brokers, the FTC found that they "obtain most of their data from other data brokers rather than directly from an original source", that seven of the nine provided data to each other, and that one obtained consumers' contact information from twenty different sources [8]. FISD's note on data lineage says "it can be challenging to sort through the agreements to identify the origin of the data and ascertain whether a particular data vendor a data buyer is dealing with has the authority to license the data set" [20].

Enforcement has turned on what vendors did and said. In 2021 the SEC brought what it called its first enforcement action charging an alternative data provider with securities fraud [1]. In 2024 the FTC's complaint charged that X-Mode Social and its successor Outlogic "failed until May 2023 to remove sensitive locations from the raw location data it sold" [4]. In 2025 the FTC alleged that Gravy Analytics and Venntel collected location data "without obtaining verifiable user consent for commercial and government uses" [6].

Who publishes the questions

FISD, the Financial Information Services Association of the Software & Information Industry Association, founded its Alternative Data Council in January 2019 [11, 19]. The council's goals include "Simplifying vendor due diligence and contracting", and its documents, "Created by a buy-side led working group", include a vendor tear sheet, due diligence questionnaires (DDQs), identification factors, service level recommendations and papers on generative models [11]. Adopting them "is entirely voluntary" [11]. They were written for investment managers that buy alternative data, and this report reads their questions as a template for any buyer of licensed data.

Banking regulators published third-party due diligence factors in 2023 [28] and proposed replacing them in September 2026 [29]. NIST's frameworks treat suppliers and third-party data as risks to govern [30, 31, 32]. Privacy law reaches a vendor's sources through California's data broker rules and privacy act [21, 22, 23], the GDPR [24] and French regulator guidance [25].

Checks on the dataset itself, its datasheet, license terms and a measured sample, are covered in buying training data (https://rillor.com/insights/buying-training-data). This report is for procurement, legal and data teams that license data from outside vendors. It is research, never legal advice.

Method

What was examined. The questions a buyer can put to a data vendor about its identity, its sources, its checks on its own suppliers, personal data, its use of generative models, its delivery terms and its conduct after onboarding, and how the answers can be verified.

Sources. Enforcement records and court opinions [1, 2, 4, 5, 6, 7, 9, 10, 34], an SEC staff risk alert [3], an FTC study [8], FISD documents [11, 12, 13, 14, 15, 16, 17, 18, 19, 20], California sources [21, 22, 23], EU and U.S. rules and French regulator guidance [24, 25, 26, 27], banking guidance [28, 29], NIST frameworks [30, 31, 32] and the AICPA's page on SOC reports [33].

How. Each document was read in full text. Complaint paragraphs are reported as FTC allegations that the respondents neither admitted nor denied [5, 7]. FISD dates come from the documents' footers, approval lines, file names or upload paths. Each legal text is summarized with its date and scope, and staff alerts, advisories and supervisory guidance are labelled as such. Questions were grouped by what they test. Web pages were viewed on 8 November 2026, and the appeals court opinion [34] on 8 November 2026.

Findings

1. Who the vendor is

FISD's Data Vendor Tear Sheet, posted in June 2021, asks "Originator or Aggregator?" [15]. The originator is "the entity that has overall responsibility for a given data item", and "A data integrator pulls in data from other sources" [15]. For an aggregator, each question about sources applies again to its suppliers.

FISD's identification factors, version 1.0 of March 2023, state that "understanding the vendor and its compliance practices may be just as important as understanding the data itself" [16]. They name tenure ("Tenure may often result in more mature compliance programs"), regulatory scrutiny or litigation, headquarters jurisdiction, officers and executives, and reputable outside counsel. The document's opening paragraph says "no single factor below is dispositive in designating a particular dataset as Alternative Data" [16].

The 2024 DDQ asks whether, in the last five years, the company and, as far as it knows, its providers have "Received a cease and desist order", "Been sued/notified of alleged claim", "Received a subpoena" or been investigated by a regulator or law enforcement agency [12]. It asks about the legal department, outside counsel and written policies, including one on "Data supplier due diligence" [12]. It asks for at least two named reviewers of the answers, one from legal or compliance where such a department exists, and for an effective date [12].

California defines a data broker as a business that knowingly collects and sells to third parties "the personal information of a consumer with whom the business does not have a direct relationship", and requires annual registration, with a 2026 fee of $6,000 plus a processing fee for electronic payments [21]. Its privacy agency's Enforcement Division advised on 17 December 2025 that "Data brokers cannot rely on a parent or affiliated entity's registration to 'cover' them", that they must list all websites and trade names, and that a broker that fails to register faces a fine of $200 for each day [22]. The advisory says it does not "constitute legal advice" [22]. A registry check covers the signing entity, its parent and each trade name.

The 2023 interagency guidance for banks lists ownership structure, "including identifying any beneficial ownership", and "whether the third party itself or any owners are subject to sanctions by the Office of Foreign Assets Control" [28]. It asks whether a third party's "statements and assertions accurately represent the activities and capabilities of the third party", and lists financial condition, key personnel and insurance [28]. It "does not have the force and effect of law" [28]. Where the Department of Justice's 2025 rule on bulk U.S. sensitive personal data applies to a restricted transaction, the U.S. person's data compliance program logs "The identity of the transaction parties, including any ownership of entities" [26].

2. Where the data comes from: originators, aggregators and lineage

The DDQ asks where data is obtained and stored, by country, state and cloud, and how it is collected, from a checklist that runs from web scraping, apps and browser plugins to third-party aggregators, connected devices, financial transactions, government and public sources, on-premise collection, surveys, and satellite

and drone collection [12]. It asks "Who provides the underlying information contained in the Data?" [12]. The tear sheet adds the "State of data on delivery (raw, semi-processed, processed)", "If processed, describe how", "Any Breaks in History?" and "Point-in-Time Methodology?" [15].

FISD's identification factors warn that "even an innocuous-seeming dataset may have heightened risk implications" when it is gathered with certain processes or technology, such as drones or mobile apps, and say firms may choose additional due diligence for data that implicates privacy concerns or comes from "restrictive jurisdictions" [16].

Lineage is the record of that path: "Data lineage identifies the source of a data set and describes the transfers, reviews and transformations applied to the data before it arrives at the customer" [20]. FISD's example questions ask whether a vendor backfills data to create a point-in-time set, interpolates missing values, changes values to fit the schema or removes outliers [20].

The EU AI Act asks providers of high-risk systems for the same record. Article 10(2) subjects their training, validation and testing data to governance practices covering "data collection processes and the origin of data, and in the case of personal data, the original purpose of the data collection", and preparation steps such as annotation, cleaning and enrichment [27].

3. How the vendor checks its own suppliers

The FTC's 2014 study found that "The majority of the data brokers in this study selects these sources based on their reputation in the industry" [8]. A few "affirmatively evaluate the legitimacy, stability, and quality of their sources", reviewing websites, terms of use, collection methods, privacy practices and regulatory compliance, and a few "ask the source for its sources" [8]. Most of the brokers put a clause in their source contracts stating that the source warrants it legally obtained the data [8]. Only two brokers required a warranty that consumers had notice of sharing and a chance to opt out, and contracts generally did not address accuracy beyond the sources' best efforts [8].

The DDQ asks the vendor to "describe any diligence you perform on the Data to determine whether the persons or companies providing the underlying information contained in the Data have the legal right to do so", and asks for redacted copies of contracts with primary data providers [12].

Two FTC orders set a program out. The X-Mode Social and Outlogic order, issued 11 April 2024, requires a Supplier Assessment Program within 90 days, "designed to ensure that consumers have provided consent for the collection and use of Location Data obtained by Respondents" [5]. The respondents must document the program in writing, assess each supplier within 30 days of a data sharing agreement or first collection "and thereafter annually", "Create and maintain records of the suppliers' responses" and stop using location data collected without consent [5]. The Gravy Analytics and Venntel order of 13 January 2025 has the same structure [7]. Both orders end 20 years after issuance, or 20 years after any later federal court complaint alleging a violation, and bind only their respondents [5, 7].

The Gravy complaint describes a program that failed, in allegations the respondents neither admitted nor denied [7]. Gravy Analytics asked suppliers for samples of consumer notices but kept using a supplier's data "even if a supplier refuses to provide this information" [7]. One supplier said its apps collect consent "[w]herever applicable regulatory rules dictate", and "Despite these ambiguous or non-responsive answers, Gravy Analytics continued to use data obtained from these sources" [7]. Venntel "relies on Gravy Analytics to obtain such confirmations" [7].

NIST's Cybersecurity Framework 2.0 calls for planning and due diligence "to reduce risks before entering into formal supplier or other third-party relationships" [30].

4. Personal data and the evidence behind it

The DDQ's appendix on Covered Personal Information asks about notice, consent and its evidence, de-identification and diligence on the vendor's own vendors, and its document list asks for "a copy of any consent or other terms" [12].

The CNIL, France's data protection authority, published guidance on re-used datasets on 7 June 2024 [25]. A re-user "must ensure that they do not re-use a dataset whose creation or transmission was manifestly unlawful", for example where its source is not indicated [25]. The CNIL recommends an agreement with the original holder stating the source, the context of collection, the legal basis, the information given to the people concerned and guarantees of lawfulness [25]. A re-user relying on consent obtained by a third party "must be able to prove that valid consent has indeed been obtained", an obligation that "cannot be fulfilled by the mere presence of a contractual clause" [25]. The contract can set how consent is demonstrated and how the evidence is kept [25].

California's privacy act, in the text effective 15 July 2024 as posted by its privacy agency in August 2024, requires a business that sells personal information to sign an agreement that "Specifies that the personal information is sold or disclosed by the business only for limited and specified purposes", among other terms [23]. A third party that bought the data may not sell or share it "unless the consumer has received explicit notice and is provided an opportunity to exercise the right to opt-out" [23]. A buyer asks whether model training sits inside the specified purposes, and whether onward licensing is allowed.

Under Article 14 of the GDPR, a controller that did not obtain data from the person tells the person, among other things, the purposes, the recipients and "from which source the personal data originate", within one month at the latest, with exceptions such as where notice "proves impossible or would involve a disproportionate effort" [24]. A buyer asks how the vendor met this duty, or which exception it relied on.

Limits also follow the data downstream. The X-Mode order requires the company to keep policies, procedures and technical measures designed to prevent recipients of its location data from using it "to determine the identity or the location of an individual's home", including contract terms that bar resale of the raw data [5].

5. How the vendor uses generative models

FISD's DDQ dated 28 February 2024 adds to the 2022 edition one question, whether the vendor uses generative tools "in the production of your Data", and an appendix on them [12, 13]. The appendix asks which tools and versions the vendor uses, public or internal, whether personal, confidential or material nonpublic information goes in, whether output is reviewed "for accuracy and/or bias", and whether the tool was trained on "data licensed from a third party" [12].

FISD's considerations of 28 January 2025 list risks including infringement, privacy, accuracy and bias, and name diligence areas for vendors: model versions, "the role that GenAI played in the production of the data product", inputs, testing and monitoring of accuracy, and the fine-tuning method [17].

FISD's March 2026 buy-side paper cites FISD's Derived Data definition, which it says "has guided industry practices since 2015": information created from the licensed data that cannot be "readily reverse-engineered to recreate the Information" or used to create "a reasonable facsimile for the Information" [18]. It tests outputs by "whether the vendor's licensed data can be reverse-engineered from the output, and whether the output data could function as a material substitute for the vendor's original data" [18].

6. Testing the answers

The App Annie order shows answers diverging from practice. The company and its founder made representations "in response to diligence questionnaires from trading firms", among other settings [2]. The order finds that "the Company did not have effective internal controls and did not conduct regular compliance reviews, and representations to the contrary were false and misleading" [2]. A policy agreed around late 2014 was not documented until April 2017, and it was amended only in June 2018, "after the Company learned of the SEC's investigation" [2]. App Annie was ordered to pay a $10 million penalty, and its founder a $300,000 penalty with a three-year bar from serving as an officer or director of a public company, without admitting or denying the findings [1].

SEC examination staff's April 2022 risk alert, which "has no legal force or effect", described advisers that did not appear to have policies to assess the terms and legal obligations of the data they bought, "including when advisers became aware of red flags about the sources of such alternative data", and advisers "that could not demonstrate, such as by producing documentation, that their policies and procedures had been consistently implemented" [3].

FISD's identification factors name checks the buyer runs itself, "reviewing information discovered by the Firm (e.g., litigation history, news articles)" [16]. Diligence documentation can include "a legal due diligence memo, record of diligence steps taken" [16]. The DDQ asks vendors to answer "after performing a reasonable investigation", and asks for a data dictionary and a sample that "should contain less than 100 rows of Data" and "should be from more than 3 months ago" [12]. The trial questionnaire, "A shorter checklist appropriate for trial data", keeps the same document list [11, 14]. Scoring that sample against the vendor's claims on labeled records follows the method of a document extraction pilot (https://rillor.com/insights/testing-document-extraction-accuracy).

The banking guidance sets the standard of proof. "Relying solely on experience with or prior knowledge of a third party is not an adequate proxy for performing appropriate due diligence" [28]. Diligence "should be commensurate with the level of risk and complexity of the third-party relationship", and an organization should "identify and document any limitations of its due diligence" [28]. The September 2026 proposal, open for comment until 16 November 2026, states that "the amount of due diligence that is warranted depends on the risk presented by the third-party relationship", and that for lower-risk relationships an organization "may be more comfortable with relying either on less-detailed due diligence information or on public or alternative sources" [29]. The agencies "plan to rescind and replace existing guidance" [29].

A vendor's methods can become the buyer's liability. Refused a Westlaw license, Ross Intelligence "made a deal with LegalEase to get training data in the form of 'Bulk Memos'", written by lawyers from a guide on using Westlaw headnotes, and bought roughly 25,000 of them to train its search tool [9]. On 11 February 2025 the trial court found "actual copying of 2,243" headnotes and wrote, "I grant summary judgment for Thomson Reuters on fair use" [9]. On 23 May 2025 it certified two questions, on originality and on fair use, for interlocutory appeal [10]. On 29 September 2026 the Third Circuit held the 2,243 headnotes "original enough for copyright protection" and Ross's use not fair, and affirmed [34].

7. Delivery, change, security and contract terms

FISD's service level recommendations, version 1.0, approved 11 June 2021, start from a delivery risk: "the delivery of the Information in production can stop abruptly when the Information Provider's source of Alternative Data is no longer available" [19]. Major changes include those that "trigger additional Compliance review due to a change in upstream provider or changing exposure to MNPI or PII", and for those the provider informs the subscriber "at least 90 days prior to implementation", with machine-readable documentation and test data that "should be same in every aspect as the updated data" [19].

The DDQ asks about "third party dependencies and/or access restrictions" such as the "expiration of upstream license agreement(s)" [12]. The tear sheet asks for update frequency, delivery delay and the "Quality Metrics Provided with Each Delivery" [15]. Where records name companies it asks "Ticker/Entity Mapping?" [15], and how entity resolution matches records (https://rillor.com/insights/how-entity-resolution-matches-records) covers how such matching is measured.

The banking guidance lists information security, operational resilience, incident reporting and "how the third party selects and oversees its subcontractors" among its factors [28]. It asks "whether the scope and the results of the SOC reports, certifications, or assessments are relevant to the activity" [28]. The AICPA's SOC 2 guide covers controls relevant to security, availability, processing integrity, confidentiality or privacy [33].

NIST's generative profile recommends contracts and service level agreements "that specify content ownership, usage rights, quality standards, security requirements, and content provenance expectations for GAI systems", and clauses that let an organization evaluate a third party's generative AI processes and

standards [32]. Where the Department of Justice rule applies, a U.S. person in covered data brokerage with a foreign person must contractually bar onward brokerage of the same data with a country of concern or covered person, and report known or suspected violations within 14 days [26].

8. After onboarding and at the end of the relationship

The FTC orders make supplier reassessment annual for their respondents [5, 7]. The banking guidance escalates issues found in monitoring, "such as material or repeat audit findings, deterioration in financial condition, security breaches, data loss, service interruptions, compliance lapses" [28]. NIST's Cybersecurity Framework 2.0 has supplier risks "monitored over the course of the relationship" (GV.SC-07) and plans for "activities that occur after the conclusion of a partnership or service agreement" (GV.SC-10) [30]. NIST's AI Risk Management Framework calls for contingency processes to handle failures or incidents in third-party data deemed high-risk (GOVERN 6.2) [31].

California sets a date for its registry: from 1 January 2029, registrants must disclose whether they have undergone the independent audit the law requires [21].

9. The questions in one table

Question	A complete answer includes	Sources
Originator or aggregator?	The originator of each data item, and an aggregator's suppliers	[8, 15]
Which entity signs?	Legal name, parent, trade names and websites, and a registry entry where a law requires one	[21, 22]
Who owns it?	Beneficial owners and a sanctions check	[26, 28]
What is its five-year record?	Orders, suits, subpoenas and regulator inquiries, for the vendor and its providers	[12, 16]
Who reviewed the answers?	Two named reviewers, one from legal or compliance, and an effective date	[12]
How was the data collected and changed?	Countries, methods, processing, backfill and interpolation	[12, 15, 20]
How are suppliers checked?	A written program, assessment dates, responses on file and what happens to vague answers	[5, 7, 8, 12]

Question	A complete answer includes	Sources
What shows the right to license?	Redacted upstream contracts with resale and derived data rights	[12]
What proves consent?	Consent records and notices, the agreed purposes and the source disclosure	[12, 23, 24, 25]
Did generative models help?	Tools, versions, inputs, output review and the tool's training data	[12, 17]
Can the answers be verified?	Data dictionary, sample, dated policies and a diligence record with its limits	[2, 3, 12, 16, 28]
What notice comes before changes?	At least 90 days for major changes, including a new upstream provider	[12, 19]
What security evidence fits?	Reports whose scope fits the service, and subcontractor oversight	[28, 33]
What happens later and at the end?	Annual reassessment, monitoring and plans for the end of the agreement	[5, 7, 28, 30]

Implications

For buyers and procurement teams. Ask the vendor-level questions before a trial, as FISD's trial questionnaire does [14], starting with originator or aggregator [15] and the legal entity that signs [21, 22].

For legal and compliance teams. Treat each answer as a claim to verify, and keep the record: documents requested, what came back and what could not be checked [3, 16, 28]. Ask for dated written policies [2]. For personal data, ask for consent evidence and the agreed purposes, and test model training against them [23, 25].

For data and model teams. Ask how the data was processed before delivery, including backfill, interpolation and generative tools [12, 17, 20], and what notice comes before a schema or upstream change [19].

For vendors. The questions above come from public documents, so any buyer can ask them. The App Annie order shows the cost of answers that practice did not support [1, 2].

Licensed datasets. Rillor's data marketplace offers ready-made datasets to license for model training and research, each stating what it holds, where it comes from and how it stays current, with use, term and onward sharing set in a written data agreement before access starts.

Limits of this analysis

- This is research, never legal advice. Each legal text is summarized with its date and scope.
- FISD's materials were written for investment managers that buy alternative data, and adopting them is voluntary [11]. No source reviewed measures how widely they are used.
- The banking guidance applies to banking organizations and has no force of law [28]. The 2026 text is a proposal [29].
- The FTC orders bind their respondents only, and the complaint paragraphs cited are allegations [5, 7].
- The Ross opinions are one trial court's rulings and the Third Circuit's affirmance on interlocutory appeal [9, 10, 34]. Later proceedings were not reviewed.
- The California privacy act is cited as posted in August 2024 [23]. Later amendments and other states' data broker laws were not reviewed, and no EU AI Act application date is stated.
- Beyond the FTC's 2014 study of nine brokers [8], no source reviewed measures how often vendors misstate their sourcing.

Sources

1. SEC Charges App Annie and its Founder with Securities Fraud, Press Release 2021-176. U.S. Securities and Exchange Commission, 14 September 2021, viewed 8 November 2026. https://www.sec.gov/newsroom/press-releases/2021-176

2. In the Matter of App Annie Inc. and Bertrand Schmitt, Order Instituting Cease-and-Desist Proceedings, Exchange Act Release No. 92975, Administrative Proceeding File No. 3-20549. U.S. Securities and Exchange Commission, 14 September 2021, viewed 8 November 2026. https://www.sec.gov/litigation/admin/2021/34-92975.pdf

3. Investment Adviser MNPI Compliance Issues, Risk Alert. U.S. Securities and Exchange Commission, Division of Examinations, 26 April 2022, viewed 8 November 2026. https://www.sec.gov/files/code-ethics-risk-alert.pdf

4. FTC Finalizes Order with X-Mode and Successor Outlogic Prohibiting it from Sharing or Selling Sensitive Location Data. U.S. Federal Trade Commission, 12 April 2024, viewed 8 November 2026. https://www.ftc.gov/news-events/news/press-releases/2024/04/ftc-finalizes-order-x-mode-successor-outlogic-prohibiting-it-sharing-or-selling-sensitive-location

5. In the Matter of X-Mode Social, Inc. and Outlogic, LLC, Decision and Order, Docket No. C-4802. U.S. Federal Trade Commission, 11 April 2024, viewed 8 November 2026. https://www.ftc.gov/system/files/ftc_gov/pdf/X-ModeSocialDecisionandOrder.pdf

6. FTC Finalizes Order Prohibiting Gravy Analytics, Venntel from Selling Sensitive Location Data. U.S. Federal Trade Commission, 14 January 2025, viewed 8 November 2026. https://www.ftc.gov/news-events/news/press-releases/2025/01/ftc-finalizes-order-prohibiting-gravy-analytics-venntel-selling-sensitive-location-data

7. In the Matter of Gravy Analytics, Inc. and Venntel, Inc., Complaint and Decision and Order, Docket No. C-4810, FTC File No. 212-3035. U.S. Federal Trade Commission, 13 January 2025, viewed 8 November 2026. https://www.ftc.gov/system/files/ftc_gov/pdf/212_3035_-_gravy_analytics_final_consent_package_without_signatures.pdf

8. Data Brokers: A Call for Transparency and Accountability. U.S. Federal Trade Commission, May 2014, viewed 8 November 2026. https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

9. Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc., No. 1:20-cv-613-SB, Memorandum Opinion. U.S. District Court for the District of Delaware, via GovInfo, 11 February 2025, viewed 8 November 2026. https://www.govinfo.gov/content/pkg/USCOURTS-ded-1_20-cv-00613/pdf/USCOURTS-ded-1_20-cv-00613-5.pdf

10. Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc., No. 1:20-cv-613-SB, Memorandum Opinion on interlocutory appeal. U.S. District Court for the District of Delaware, via GovInfo, 23 May 2025, viewed 8 November 2026. https://www.govinfo.gov/content/pkg/USCOURTS-ded-1_20-cv-00613/pdf/USCOURTS-ded-1_20-cv-00613-6.pdf

11. Alternative Data Council. FISD, Software & Information Industry Association, undated page, viewed 8 November 2026. https://fisd.net/alternative-data-council/

12. Data Provider Due Diligence Questionnaire ("DDQ"), Compliance Due Diligence Questionnaire Including Generative AI Questions. FISD Alternative Data Council, 28 February 2024, viewed 8 November 2026. https://fisd.net/wp-content/uploads/2024/02/FISD-Alternative-Data-Council-Due-Diligence-Questionnaire-with-GenAI-Questions-

022824.docx

13. Data Provider Due Diligence Questionnaire ("DDQ"). FISD Alternative Data Council, 19 May 2022, viewed 8 November 2026. https://fisd.net/wp-content/uploads/2022/05/Alternative-Data-Council-Due-Diligence-Questionnaire-051922.docx

14. Data Provider Trial Due Diligence Questionnaire ("DDQ"). FISD Alternative Data Council, 19 May 2022, viewed 8 November 2026. https://fisd.net/wp-content/uploads/2022/05/Alternative-Data-Council-Trial-Due-Diligence-Questionnaire-051922.docx

15. Data Vendor Tear Sheet. FISD Alternative Data Council, June 2021, viewed 8 November 2026. https://fisd.net/wp-content/uploads/2021/06/Data-Vendor-Tear-Sheet.docx

16. Alternative Data Identification Factors, Version 1.0. FISD Alternative Data Council, Software & Information Industry Association, March 2023, viewed 8 November 2026. https://fisd.net/wp-content/uploads/2023/03/FISD-Alternative-Data-Identification-Factors-11.1.22.pdf

17. FISD Alternative Data Council: Compliance Considerations for Generative AI, v 1. FISD Alternative Data Council, 28 January 2025, viewed 8 November 2026. https://fisd.net/wp-content/uploads/2025/01/FISD-Alternative-Data-Council-Compliance-Considerations-for-Generative-AI-012825.pdf

18. FISD Alternative Data Council Buy-Side Perspective: Understanding Investment Firms' Use of Licensed Data with GenAI, Version 1.0. FISD Alternative Data Council, March 2026, viewed 8 November 2026. https://fisd.net/wp-content/uploads/2026/03/FISD-Alternative-Data-Council-Buy-Side-Perspective-Understanding-Investment-Firms-Use-of-Licensed-Data-with-GenAI-1.pdf

19. Best Practice Recommendations on Alternative Data Service Levels, Response Times and Communications Procedures, Version 1.0. FISD, Software & Information Industry Association, approved 11 June 2021, viewed 8 November 2026. https://fisd.net/wp-content/uploads/2021/09/FISD-Best-Practice-Recommendations-on-Alternative-Data-Service-Levels-Response-Times-and-Communications-Procedures-v-01.0-June-2021.docx

20. Alternative Data Standards Learning Point: Data Lineage. FISD Alternative Data Council, June 2021, viewed 8 November 2026. https://fisd.net/wp-content/uploads/2021/06/Data-Lineage-Learning-Point.pdf

21. Information for Data Brokers. California Privacy Protection Agency, undated page, viewed 8 November 2026. https://cppa.ca.gov/data_brokers/

22. Enforcement Advisory No. 2025-01, Data Broker Registration: Trade Names, Websites, and Parents/Subsidiaries. California Privacy Protection Agency, Enforcement Division, 17 December 2025, viewed 8 November 2026. https://cppa.ca.gov/pdf/enfadvisory202501.pdf

23. California Consumer Privacy Act of 2018, Cal. Civ. Code § 1798.100 et seq., text effective 15 July 2024 (AB 3286 update). California Privacy Protection Agency, August 2024, viewed 8 November 2026. https://cppa.ca.gov/regulations/pdf/cppa_act.pdf

24. Regulation (EU) 2016/679 (General Data Protection Regulation). European Parliament and Council of the European Union, EUR-Lex, 27 April 2016, viewed 8 November 2026. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679

25. Ensuring the lawfulness of the data processing: In case of re-use of data, carrying out the necessary additional tests and verifications. Commission nationale de l'informatique et des libertés (CNIL), 7 June 2024, viewed 8 November 2026. https://www.cnil.fr/en/ensuring-lawfulness-data-processing-re-use

26. Preventing Access to U.S. Sensitive Personal Data and Government-Related Data by Countries of Concern or Covered Persons, 28 CFR Part 202. U.S. Department of Justice, final rule published 8 January 2025 (90 FR 1636), eCFR text current as of 5 November 2026, viewed 8 November 2026. https://www.ecfr.gov/current/title-28/chapter-I/part-202

27. Regulation (EU) 2024/1689 (Artificial Intelligence Act). European Parliament and Council of the European Union, Official Journal of the European Union, 13 June 2024, published 12 July 2024, viewed 8 November 2026. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689

28. Interagency Guidance on Third-Party Relationships: Risk Management, 88 FR 37920. Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation and Office of the Comptroller of the Currency, 9 June 2023, viewed 8 November 2026. https://www.federalregister.gov/documents/2023/06/09/2023-12340/interagency-guidance-on-third-party-relationships-risk-management

29. Proposed Third-Party Risk Management Guidance, 91 FR 58536. Office of the Comptroller of the Currency, Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation and National Credit Union Administration, 15 September 2026, viewed 8 November 2026. https://www.federalregister.gov/documents/2026/09/15/2026-18859/proposed-third-party-risk-management-guidance

30. The NIST Cybersecurity Framework (CSF) 2.0, NIST CSWP 29. National Institute of Standards and Technology, 26 February 2024, viewed 8 November 2026. https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf

31. Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1. National Institute of Standards and Technology, January 2023, viewed 8 November 2026. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

32. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, NIST AI 600-1. National Institute of Standards and Technology, July 2024, viewed 8 November 2026. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

33. System and Organization Controls (SOC) suite of services. AICPA & CIMA, undated page, viewed 8 November 2026. https://www.aicpa-cima.com/resources/landing/system-and-organization-controls-soc-suite-of-services

34. Thomson Reuters Enterprise Centre GmbH v. ROSS Intelligence Inc., No. 25-2153, Opinion of the Court. U.S. Court of Appeals for the Third Circuit, filed 29 September 2026, viewed 8 November 2026. https://www2.ca3.uscourts.gov/opinarch/252153p.pdf

About this report

Published by Rillor on 9 November 2026. Superintelligence (SI) tools assist research and drafting, and every figure cites its source. Rillor builds and secures SI systems, supplies the data they learn from and delivers the infrastructure they run on. We apply SI to financial and compute markets.

Cite as: Rillor. Due diligence questions to ask a data vendor. 9 November 2026. https://rillor.com/insights/data-vendor-due-diligence-questions

Work with Rillor

Send the datasets you need, the coverage period, the delivery format and the intended use. Contact us at rillor.com/contact?type=market-data or contact@rillor.com. More on data marketplace: rillor.com/data/marketplace