

Data Provider Due Diligence Questionnaire (“DDQ”)

Because investment managers intend to use this data to inform investment decisions for trading public and private securities, they must consider applicable securities laws, as well as other laws, in evaluating data sets. The responses you provide here will likely be material to an investment manager's decision to acquire and use the data product(s).

As a general rule, investment managers do not want to receive data containing and/or derived from the following types of information:

- **Material, non-public information (also known as “MNPI” or “Inside Information”)** (i.e., information that is not widely known and could affect a reasonable investor's decision to trade);
- **Information that is subject to a confidentiality obligation;**
- **Information that the vendor is prohibited from disclosing;**
- **Information obtained through misappropriation, through deception, in violation of law, or in breach of a duty (whether contractual, fiduciary, or otherwise);** and
- **Personal Information, Personally Identifiable Information and Personal Data (referred to herein as “Covered Personal Information”)** (i.e., information that could be used to identify or be associated with individual natural persons).

DEFINITIONS

The following definitions apply:

- “Company” means the name of the entity offering the Data.
- “Affiliate” means any entity that controls, is controlled by, or is under common control with the Company.
- “Data” means the information product offered by the Company and, where appropriate, any information used directly or indirectly to create, process, structure, acquire, aggregate, or resell the Data.
 - When the words “information,” “data,” or “data set” are used, without having been capitalized, please interpret them according to their ordinary dictionary meaning.
- “MNPI” and “Inside Information” have the meanings used above.
- “Covered Personal Information” has the meaning used above.

INSTRUCTIONS

Please answer each question in this DDQ, responding with all information available to you and which you believe to be true after performing a reasonable investigation. If you must make an interpretation of any question or any term used within a question, please describe that interpretation as part of your answer.

DOCUMENT REQUEST LIST

In addition to answering the questions below, please provide the following, if they exist for the Data:

- A document describing the Data, such as a tear sheet, brochure, or other marketing literature
- A data dictionary, data schema, or other document listing the fields provided in the Data and describing the types of values that may be contained within each field
- A small sample of the Data in human-readable form. This sample should contain less than 100 rows of Data. The sample Data should be from more than 3 months ago
- If you buy Data from others, a copy of any relevant terms from your contracts with those persons. These terms should show your legal right to resell or license the Data to us and/or to sell/license newly created (i.e. “derived”) Data products
- If you collect or generate Data from or about individuals or their devices, a copy of any consent or other terms you have entered into with those individuals concerning Data collected from or about them or their devices

1 COMPANY INFORMATION

1.1	What is the Company's name?	Nosible Inc
1.2	Does the Company have a different legal name?	No. Nosible Inc is our full legal name. We trade as NOSIBLE.
1.3	Company's principal business/headquarters address:	1209 Orange Street, Wilmington, New Castle County, Delaware, 19801, United States
1.4	Please provide the registered address of the Company if different from its principal business address.	Same as the principal business address.
1.5	Years in business:	The business was founded in the United Kingdom, where NOSIBLE Ltd was incorporated on September 15, 2021 (company number 13623107). Nosible Inc was incorporated in Delaware on October 24, 2024 when the business moved to the US. Nosible Inc is the entity offering the Data.
1.6	Names of any Affiliates, if relevant with respect to the Data:	NOSIBLE Ltd, a UK company (company number 13623107). The business operated through NOSIBLE Ltd before moving to the US in 2024.
1.7	Is Company or any Affiliate a regulated entity?	☐ Yes, and details are provided below ☒ No Not applicable.
1.8	Do you model, maintain a "paper book" or otherwise track potential performance of your data product(s) against the public equity markets? If yes, do you publish or share with clients?	No. We do not model or track the potential performance of the Data against public equity markets, and we do not publish or share any performance track record.
1.9	Please describe any formal or informal investment policies the Company maintains concerning its investments or the investments of its employees.	We are a technology company, not an investment firm. NOSIBLE does not make investments based on the Data. There are no restrictions on employee investments at present.
1.10	Does the Company make investments or allow its employees to make investments based on the Data?	☐ Yes, and details are provided below ☒ No Not applicable.

1.11	What is the Company's website address?	https://www.nosible.com
1.12	Name of Executive Relationship Manager, Account Representative, or other primary sales contact:	Stuart Reid, CEO and Founder. stuart@nosible.com, +1 (646) 891-1234.
1.13	Please provide the names of at least two persons who have reviewed the responses to this questionnaire and what function they hold at the Company. If you have a legal or compliance department, at least one person from one of those departments should review these responses.	Stuart Reid (stuart@nosible.com), CEO and Founder. Matthew Dicks (matthew@nosible.com), Founding Engineer.
1.14	Effective date of this questionnaire:	July 10, 2026

2 Product/Service Information

2.1 GENERAL

2.1.1	What is the name of the Data? Please provide any alternate names that Data may be known by.	NOSIBLE World. Also referred to as the NOSIBLE World database. Version 1.2 is current; version 1.1 is retained for continuity.
2.1.2	Please describe the Data or provide a description from your data catalogue, tear sheet, or other documentation.	NOSIBLE World is a point-in-time database of global events built from NOSIBLE's web index. It contains one slice per day from January 1, 2015 through July 1, 2026. Each event carries a title, description, date, language, geography, sentiment and materiality signals, coverage statistics, entities, tickers and security identifiers, classifications across several taxonomies (IAB, IPTC, GICS, asset class, SDGs, and the NOSIBLE taxonomy), provenance records with source URLs and snippets, and an embedding. The Data is delivered as four artifacts per day via a private S3 bucket with per-customer read-only access keys. NOSIBLE World is capable of being deployed on-prem in a privacy preserving and secure way.
2.1.3	In what year did the Company begin offering the Data to customers?	2026. NOSIBLE World was first offered to customers in June 2026.
2.1.4	Approximate number of active customers in each category below that the Company provides the Data to: - Financial services related customers? - Non-Financial services related customers?	Approximate number of financial services related customers: 1-10 Approximate number of non-financial services related customers: 1-10
2.1.5	Is the Data disseminated simultaneously to all customers?	☒ Yes ☐ No, and explanation is provided below Not applicable.

2.1.6	Are there any third party dependencies and/or access restrictions to the Data that would impact our receipt and/or use of the Data (e.g., expiration of upstream license agreement(s), anticipated updates to website/app terms of service through which the data is collected, changing privacy regulations (i.e., CPRA)?	☒ Yes, and explanation is provided below ☐ No Semantic search against the shipped vector index requires query embeddings from OpenAI's text-embedding-3-large model. Filtering, sorting, and lookups do not require any third party. Upstream, our crawler routes most requests through Bright Data and we are evaluating fallback providers. We do not license the underlying data from anyone, so there are no upstream license agreements that could expire.
2.1.7	Do you collect or generate the Data yourself, or do you buy or collect the Data from others? If you either buy or collect the Data from others, please provide any relevant Terms & Conditions or Agreements with those persons related to the Data, such as user consents to collection and dissemination of anonymized personal data or terms demonstrating your ability to resell the Data or to create derived Data products. You may redact such agreements as necessary to comply with your confidentiality obligations.	We generate the Data ourselves. Every page behind NOSIBLE World was crawled from the public web by our own systems. We do not buy or license the underlying data from anyone, so there are no third-party terms to provide.

2.2 DATASET INFORMATION

2.2.1	From which countries/states/clouds are the Data obtained?	The Data is obtained from publicly available websites in more than 150 countries. We crawl public written sources only: news, press releases, company websites, specialist publications, government pages, public records, official filings, and court documents.
2.2.2	Where are the Data stored? In which countries/states/clouds are the Data stored?	Primary indexes run on dedicated OVHcloud bare-metal servers in France, Germany, and the UK. Raw HTML and parsed JSON are stored in S3 buckets hosted by Wasabi. The NOSIBLE World distribution is delivered from a private S3 bucket in us-east-1. Indexes are backed up to S3 and geographically replicated. NOSIBLE World is capable of being deployed on-prem in a privacy preserving and secure way.
2.2.3	Please indicate whether any of these data generating or collection methods are used to create the Data:	☒ Web scraping ☐ Apps; browser plugins; from consumers generally ☐ From third party aggregators ☐ IoT devices ☐ Financial transaction/exchange ☒ Government/public sources (please explain) ☐ On-premise data collection (e.g. counting cars in parking lots) ☐ Calls with public companies, expert networks, surveys of companies/individuals ☐ Satellite/drone/radar collection
2.2.4	Do you use Generative AI tools (e.g., ChatGPT) in the production of your Data? "Generative AI" means generative pre-trained transformers and large language models that produce new content in response to a user's prompts based on the data the models were trained on. If yes, please fill out Appendix 4, which refers to "Generative AI".	Yes. We use large language models in the production of the Data. Appendix 4 is completed below.

2.2.5	Who provides the underlying information contained in the Data?	The publishers of public web pages: news websites, newswires, company websites, specialist publications, government websites, public records, and court documents. The index draws from more than 300,000 vetted sources in over 150 countries and 95 languages.
2.2.6	Please 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?	Every domain must be traceable to a legal entity and pass AI vetting before it enters the source universe. Before approval we record homepage response metadata, redirect status, blocklist status, Google index presence, Google Web Risk status, RDAP registration signals, robots.txt, ads.txt, security.txt, legal-document paths, content sitemaps, and archive provenance. We look for ownership clues in homepage footers, copyright strings, privacy policies, and terms of service, and extract the legal entity name and jurisdiction where the evidence is available.
2.2.7	Please describe any other due diligence performed on third-party data suppliers ?	We do not buy data from third-party suppliers. For infrastructure vendors, we require that proxy networks source IPs legally and ethically at scale, with clear sourcing. We maintain lists of alternative vendors for each provider so we can move if needed.
2.2.8	Does the company have contracts in place with primary data providers? If yes, please provide redacted copies of those contracts.	☐ Yes, and the contracts are provided as described below ☒ No Not applicable. We have no primary data providers. The Data is built from public web pages crawled by our own systems.

3 MNPI, Covered Personal Information, and Other Issues

3.1 MNPI/INSIDE INFORMATION AND COVERED PERSONAL INFORMATION

3.1.1	Does the Data contain any MNPI or inside information, or is it potentially derived from MNPI or inside information?	☐ Yes, and more details are provided below ☒ No The Data is built entirely from publicly available written content. We do not crawl behind paywalls, logins, or other access controls.
3.1.2	Does the Company engage in any activities that involve, or expose it to, MNPI, regardless of whether the Data relating to this request contains or is potentially derived from MNPI?	☐ Yes, and more details are provided below ☒ No Not applicable. We crawl public web pages only.
3.1.3	Does the Data contain or is it derived from any information that is otherwise confidential? For example, your license to the Data is covered by a non-disclosure agreement.	☐ Yes, and more details are provided below ☒ No Not applicable. We do not license the underlying data and it is not covered by any non-disclosure agreement.
3.1.4	Please describe any diligence you perform on the Data concerning issues surrounding MNPI, inside information, or otherwise confidential information?	Our crawl scope removes the risk at the source. We crawl publicly available, logged-out pages only, with no paywall, login, or access-control bypass. Excluded source types are screened out before a page enters the index, and every domain must pass verification checks before it becomes a production source. Because we take nothing from gated systems, the Data is not exposed to MNPI, inside information, or confidential information.

3.1.5	Have you used reasonable best efforts to determine that any data you use to create the Data was not derived from a breach of fiduciary duty, breach of contract or deceptive behavior? And are you reasonably sure the Data does not contain any of the foregoing?	Yes. All data is crawled from public, logged-out web pages. We do not bypass access controls, evade blocks, or continue after individualized revocation. We are reasonably sure the Data does not contain any of the foregoing.
3.2 PERSONAL DATA / PII		
3.2.1	Does the Company, or do the third party sources of any underlying Data used by the Company to produce the Data, collect any Covered Personal Information as part of its products and/or services? If Yes, please also answer the questions in Appendix 1	☐ Yes, and more details are provided below ☒ No We do not index personally identifiable information and hold no personal data in the search index. Pages containing personally identifiable information are excluded before indexing. Where public-source text still names individuals, we handle it through minimization, source exclusions, public privacy information, and routes for objection, delisting, and removal. NOSIBLE World is capable of being deployed on-prem in a privacy preserving and secure way.
3.2.2	How does the Company identify whether it, or the third party sources of any underlying data used by the Company to produce the Data, collects any Covered Personal Information?	The crawler's exclusion policy removes the categories that create personal-data risk before content reaches the index: pages containing personally identifiable information, social platforms, marketplaces, unsafe domains, and inherently pornographic sites. Domains are screened with Google Safe Browsing and Google Web Risk, checked against blocklists, and vetted by AI before entering the source universe. We have no third-party data sources.

3.3 OTHER ISSUES

3.3.1	Does the Company collect any Data through web scraping or other automated data harvesting (e.g., does the Company deep-link, scrape, crawl or use robots, spiders, or other automated programs, devices, algorithms or methods to collect Data from websites)? If yes, please fill out Appendix 2, which refers to these practices as “scraping.”	☒ Yes ☐ No
3.3.2	Does the Company collect any Data through the use of Application Programming Interface (“API”) calls or similar techniques? If yes, please fill out Appendix 2, which refers to these practices as “scraping.”	☐ Yes ☒ No
3.3.3	Does the Company collect Data from mobile devices, such as phones or tablets or collect data about individuals’ digital device usage? If yes, please fill out Appendix 3.	☐ Yes ☒ No

4 COMPANY LEGAL AND REGULATORY INFORMATION

4.1 COMPANY’S INTERNAL LEGAL AND REGULATORY CONTROLS

4.1.1	Does the Company have a legal department and, if so, how many employees make up the function?	☐ Yes, the number of employees is provided below ☒ No Not applicable.
4.1.2	Please provide the name of the Company’s outside counsel, if any, relating to the Data and/or the Company’s data practices more generally.	NEXT Legal (United States).

4.1.3	Please provide a description of how the legal and compliance aspects of the collection and dissemination of data are managed by the Company.	Crawl scope is defined before a page enters the index. Every domain must be traceable to a legal entity and pass AI vetting, and excluded source types (paywalls, logged-in pages, social platforms, marketplaces, unsafe domains, and pages containing personally identifiable information) are screened out upfront. We publish our positions on crawling, copyright, privacy, and security at nosible.com/legal. We honor robots.txt, machine-readable rights reservations, and source-removal requests. Retrieval endpoints are rate-limited and monitored. We have completed SOC 2 readiness work with Vanta and Cognisys and are preparing for an independent SOC 2 examination. NOSIBLE World is capable of being deployed on-prem in a privacy preserving and secure way.
-------	--	--

4.1.4	Does the Company maintain written policies and procedures regarding the handling sensitive information, including: • MNPI • Covered Personal Information • Confidential information • Data supplier due diligence, if applicable?	☒ Yes, a description of the policies is provided below ☐ No Our policies covering data handling, source exclusions, privacy, and security were implemented as part of our SOC 2 readiness work, and our public postures on crawling, copyright, privacy, and security are published at nosible.com/legal. MNPI and confidential information are addressed through crawl scope: public pages only, with no access-control bypass. Covered Personal Information is addressed through the crawler's exclusion policy. Data supplier due diligence is not applicable because we do not buy data.
4.2 GENERAL DATA COLLECTION AND DATA DISSEMINATION		
4.2.1	How does the Company ensure it complies with the terms of use of any data source from which data is obtained (e.g., websites, third-party suppliers)?	We honor robots.txt and machine-readable rights reservations. We do not crawl pages that require affirmative assent to terms, such as logins, click-to-accept gates, or paywalls. During website verification we record robots.txt, terms links, privacy policy links, and legal-document paths for each source. Source owners can request removal at any time and we honor those requests.

4.2.2	Relating to its data collection/use/distribution practices, in the last 5 years, has the Company ever? Please provide additional details, if necessary	☐ Received a cease and desist order ☐ Been sued/notified of alleged claim ☐ Received a subpoena ☐ Been subject to an investigation by a regulator, law enforcement agency, or governmental organization ☐ Other ☒ None of the above None.
4.2.3	With respect to the providers of underlying data used by the Company, and their collection/ use/ distribution practices, is the Company aware whether in the past 5 years, any of those provider have ever?	☐ Received a cease and desist order ☒ Been sued/notified of alleged claim ☐ Received a subpoena ☐ Been subject to an investigation by a regulator, law enforcement agency, or governmental organization ☐ Other ☐ None of the above Bright Data, our proxy infrastructure provider, has been involved in litigation over web scraping, including Meta v. Bright Data, in which the court rejected Meta's contract claims against logged-out public scraping in January 2024. The providers of the underlying content are the publishers of public web pages; we are not aware of any relevant proceedings against them in connection with our use.

Appendices

Please answer the questions in one or more of the appendices below if you were directed to do so by one of the questions in the main body of the questionnaire.

1 COVERED PERSONAL INFORMATION		
1.1	Are natural persons provided notice regarding the commercialization of their Covered Personal Information? (either by the Company or by the third-party sources of any data used to produce the Data)?	☐ Yes ☐ No ☒ Not applicable
1.2	Do natural persons provide consent to the commercialization of their Covered Personal Information?	☐ Yes ☐ No ☒ Not applicable
1.3	Please provide copies/links of any relevant, related evidence regarding notice provided to or consent received from natural persons to the Company itself and/or to any third party sources of any underlying data used to produce the Data.	Not applicable. The Data contains no Covered Personal Information.
1.4	Does the Company or any third party sources used to produce the Data employ any de-identification or anonymization processes?	☐ Yes, and more details are provided below ☐ No ☒ Not applicable Not applicable.

1.5	Please describe any diligence you perform on the Data concerning issues surrounding Covered Personal Information or other data protected by privacy laws. This includes not only the Data, but also third party sources of any underlying data, as well as the underlying data provided by those third party sources. In other words, please describe diligence you perform on your data vendors (if any) and the data they provide.	We do not use data vendors. The Data is built from public web pages crawled by our own systems. Personal-data risk is handled before content reaches the index: pages containing personally identifiable information are excluded, along with social platforms, marketplaces, unsafe domains, and inherently pornographic sites. Individuals can raise objections, erasure requests, or delisting requests, and websites can ask to be forgotten, in which case we remove their data from the index and notify affected paying customers. NOSIBLE World is capable of being deployed on-prem in a privacy preserving and secure way.
-----	---	---

2 WEB SCRAPING AND APIs

2.1	Do you use any service providers to assist you in web data collection including proxies, CAPTCHA solutions, third-party scraping providers, or others? <u>IF SO, YOUR ANSWERS BELOW INCLUDE BOTH YOUR PRACTICES AS WELL AS THOSE OF THE ABOVE-REFERENCED SERVICE PROVIDERS</u>	☒ Proxies ☐ CAPTCHA solutions ☐ Third-party scraping providers ☐ Others
2.2	Has the Company considered any industry or other web data collection guidelines including - The FISD Web Data Collection Guidelines - Investment Data Standards Organization Web Crawling Best Practices - Others? Please describe how you implement and/or comply with that guidance.	We maintain a published crawling posture at nosible.com/legal/crawling that covers the same ground as the FISD Web Data Collection Guidelines: honor robots.txt, no access-control bypass, no block evasion, reasonable crawl rates, cache-aware requests using ETags and Last-Modified dates, and vetting of every domain before it enters the index. We have not formally certified against a specific industry framework.

2.3	Does the Company perform any analysis of the terms and conditions of the websites that it scrapes? If so, please describe that analysis.	☒ Yes, and details are provided below ☐ No Website verification records robots.txt, terms-of-service links, privacy policy links, and legal-document paths before a domain is approved. We crawl public, logged-out pages only, so we never assent to click-to-accept terms. We avoid the conduct at issue in hiQ v. LinkedIn, Facebook v. Power Ventures, and Craigslist v. 3Taps: no access-control bypass, no block evasion, and no continuing after individualized revocation.
2.4	Does the Company scrape data from directories or files specified as disallowed in robots.txt?	☐ Yes ☒ No
2.5	Does the Company utilize proxies, including anonymizing proxies? If so, do the proxy providers include information sufficient for scraped websites to contact the Company?	☒ Yes, and details are provided below ☐ No We own thousands of dedicated IP addresses and try to access pages from the country where the website operates before using more advanced proxies from Bright Data. Like other search engines, our crawler does not advertise itself in its user agent; this ensures reliable access to public pages while we maintain responsible crawl rates. Proxies are used for reliable public-page access and routing, not to evade individualized blocks, access controls, or revocation notices. NOSIBLE is publicly contactable through nosible.com and we honor removal requests.
2.6	Does the Company use multiple IP addresses or otherwise anonymize its IP addresses?	☒ Yes, and details are provided below ☐ No We own thousands of dedicated IP addresses around the world and route most requests through Bright Data. IP diversity is used for reliable access and country-aware routing, not to evade individualized blocks or revocation.
2.7	Does the Company use onion routing in connection with its scraping activities?	☐ Yes, and details are provided below ☒ No Not applicable.

2.8	Does the Company collect data from websites that require an affirmative action to access the data (e.g., checking a box to accept its terms and conditions, providing a user identification and/or password, completing a CAPTCHA or completing a similar verification mechanism)?	☐ Yes, and details are provided below ☒ No No. We do not crawl behind paywalls, logins, CAPTCHAs, click-to-accept terms, or any other access controls. When we encounter a CAPTCHA we do not solve it; we request the page again through our proxy provider.
2.9	Are the websites from which the Company scrapes able to identify and contact the Company?	☐ Yes ☒ No Our crawler does not identify itself in its user agent, so websites cannot identify NOSIBLE from crawl traffic alone. NOSIBLE is publicly contactable through nosible.com, we honor robots.txt, and source owners can request removal from the index at any time.

3 DIGITAL DEVICE/MOBILE DATA

3.1	Does the Company monitor, or is the Data or underlying data obtained by monitoring, the browsing habits of individuals?	☐ Yes, and details are provided below ☒ No Not applicable.
3.2	Does the Company process, or is the Data or underlying data obtained by processing, geolocation data from individuals' digital devices?	☐ Yes, and details are provided below ☒ No Not applicable.
3.3	Does the Company process, or is the Data or underlying data obtained by processing, data from Apple or Android mobile devices? If so, please provide details of steps the Company takes to ensure its practices conform to development guidelines and/or the legal requirements of the relevant application platform.	☐ Apple device data ☐ Android device data Not applicable. We do not collect data from mobile devices or data about individuals' digital device usage.

4 GENERATIVE AI

4.1	What is/are the name(s) of the Generative AI tool(s) you use in the production of your Data?	Large language models accessed through the OpenAI API.
4.2	Is/are the Generative AI tool(s) the public version of such tool(s) (which, for clarity, includes any public version of a Generative AI tool as well as any other version of a Generative AI tool (even if not public) where your inputs could be shared with third parties or used to train the model), or the internal version of such tool(s)?	We use the API versions, not the public consumer tools. Under OpenAI's API terms our inputs are not used to train their models. In any case, our inputs are publicly available web content.
4.3	How are Generative AI tool(s) used in the production of your Data?	Large language models are used to vet domains before they enter the source universe, to generate website and shard metadata such as titles and descriptions, and to structure crawled content into the event records that make up NOSIBLE World. All inputs are publicly available web content crawled by us.
4.4	Do you input into the Generative AI tool(s) any: PII, PHI, MNPI, confidential information, or intellectual property-protected information? If yes: Please explain, and please also provide information relating to any controls in place around those inputs.	No PII, PHI, MNPI, or confidential information. Inputs are publicly available web content, which may be protected by copyright. Our use is non-substitutive: we produce indexes, metadata, classifications, and short attributed snippets, not replacements for the original works. Retrieval endpoints are rate-limited and monitored to prevent reconstruction of complete works.
4.5	Do you review the output for accuracy and/or bias, as applicable? If yes: Please describe and provide any related policies or procedures.	Yes. Published dates are verified before entering research use through metadata checks, temporal falsification tests, corroborating sources, authoritative systems of record, archive evidence, and sensibility checks. Every step in our query and production pipeline must improve performance on our evals to stay in the system.
4.6	Please describe how the Generative AI tool was trained, where the training data set came from and how it works.	We use third-party foundation models. We do not train them. The vendors train these models on their own data and publish the details. We do not fine-tune models on customer data.

4.7	Are the Generative AI tools trained on any: (a) customer data and/or any other, underlying data subject's or end user's data? (b) data licensed from a third party? If yes: Please explain, including details of any diligence performed on the training data to determine whether the persons or companies providing the underlying information training the Generative AI tool have the legal right to do so. To the extent applicable, please provide a copy of any disclosures/consents/licensing language from the relevant contract, arrangement or terms of service (subject to reasonable redaction of sensitive and irrelevant information) related to such data.	No. We do not train or fine-tune Generative AI tools on customer data or on data licensed from third parties. The models we use are trained by their vendors on their own data.
-----	---	--