Revisiting Online Data Markets in 2022

A Seller and Buyer Perspective

Javen Kennedy, Pranav Subramaniam, Sainyam Galhotra, Raul Castro Fernandez
University of Chicago
{kennedyjaven, psubramaniam, sainyam, raulcf}@uchicago.edu

ABSTRACT

Well-functioning data markets match sellers with buyers to allocate data effectively. Although most of today's data markets fall short of this ideal, there is a renewed interest in online data marketplaces that may fulfill the promise of data markets. In this paper, we survey participants of some of the most common data marketplaces to understand the platforms' upsides and downsides. We find that buyers and sellers spend the majority of their time and effort in price negotiations. Although the markets work as an effective storefront that lets buyers find useful data fast, the high transaction costs required to negotiate price and circumvent the information asymmetry that exists between buyers and sellers indicates that today's marketplaces are still far from offering an effective solution to data trading. We draw on the results of the interviews to present potential opportunities for improvement and future research.

1. INTRODUCTION

In theory, *data markets* hold the promise of creating value in an increasingly data-driven economy by enabling transactions for sellers who own data that is useful to buyers. In practice, they do not always fulfill their potential. Some data markets are ubiquitous, with billions of participants trading data daily, such as the barter market formed by online services that offer search, social networks [5, 7] and other services to individuals in exchange for their data [1, 10]. Other markets are opaque, such as the data broker industry [19, 45, 46] that coordinates business-to-business data exchange, sometimes trading data about subjects who are not aware of the transaction. In barter and broker data markets, the market controller often reaps most of the benefits, sometimes at the expense of privacy of the involved subjects.

A promising direction to increase the value of data for more participants is *online data marketplaces*, which are seeing a renewed interest. These online platforms such as AWS Marketplaces [22] and Snowflake [51] provide sellers pricing options for their datasets, and provide buyers search functionality to identify useful datasets. These marketplaces promise to allow data-hungry organizations to extract value from datasets offered for sale.

In this paper, we study how these online data marketplaces work, and assess how closely they meet the objective of efficient allocation of valuable data. To do so, we conduct a series of structured interviews with market participants that shed light on their perspective on the upsides and downsides of today's marketplaces. We analyze the data we gathered after surveying around 70 participants and conclude with a series of challenges, including in data management, that we believe must be tackled for these markets to fulfill their promise.

We conclude that modern online data marketplaces are still far from reaching their potential. They act as a storefront, where sellers can list their offerings and buyers have a central place to identify interesting data. As a storefront, marketplaces reduce the search cost for buyers, and help them quickly identify valuable datasets. But they fall short on the promise of distributing the value of data widely and efficiently, as transaction costs remain high. This is mainly because every transaction requires a one-on-one negotiation: buyers want to make sure the data they are buying is useful and sellers want to avoid releasing the data before obtaining a payment. Today's marketplaces do not incorporate any feature to assist with resolving this information asymmetry, which is fundamental to trading data [21]. Furthermore, current marketplaces lack support to help sellers document and publish datasets and identify competitive prices.

Our work revisits the results presented in [50, 44] a decade ago. We share with that previous work the intention of understanding the data marketplace landscape. But we focus on the barriers found by market participants and conduct our work using data marketplaces that did not exist a decade ago. In fact, the marketplaces analyzed in [50, 44] have now largely disappeared. In addition, our work complements a recent survey [23] that presents an in-depth study of the kinds of organizations that trade data based on a thorough online data collection effort. Complementary to this work, we interview and survey those entities, face-to-face, to understand the barriers in trading data. Then, we use this information to propose directions for improvement. To recruit participants, we took advantage of the relatively recent introduction of the California Consumer Privacy Act (CCPA) [3], which requires any data broker to list their name and company information online [46].

In Section 2, we overview different data markets and explain how data is traded in data marketplaces in Section 3. We then present the user study in Section 4, a list of improvement opportunities for marketplaces in Section 5 and the conclusion in Section 6

2. LANDSCAPE OF DATA MARKETS

Definition. In its simplest form, a data market consists of sellers who have data, buyers who want data, and sometimes an intermediary that coordinates transactions between sellers and buyers. Data may be exchanged directly, by sharing raw datasets, or indirectly, by services provided on top of data, such as queries [25, 24, 40, 42], machine learning and statistical inferences [20, 34, 39]. For a classification of data markets, we refer the reader to [41, 49] where the authors present a taxonomy.

2.1 Market Types

The B2B data broker industry. We refer to data brokers as organizations that collect, generate, or pool data from diverse sources and resell it to interested buyers. Examples include Acxiom [19], Nielsen [45], and others, such as those listed in California's Data Broker registry [46]. In general, data brokers sell data via a sharing agreement that determines what the buyer receives. Although brokers can sell any kind of data, some recognizable businesses dominate sales. For example, *customer segmentation* data permits companies to classify customers in different groups and is hence used for marketing purposes. Another example is financial data, as sold by Bloomberg [2] and Morningstar [8] to traders and other interested parties. Data brokers are private businesses and data transactions are typically opaque, making it difficult to derive much insight about brokers' pricing strategies.

Data bartering – Social networks and Individual's Data. Individuals trade their personal data in exchange for services such as email [6], social networks [5, 7], entertainment [10], online shopping [1], and more. We mention these data markets because the exchanged good is data, and companies make profit from access to that data via different revenue streams. For example, they sell advertisements, and offer users content recommendations that increase the time individuals spend in the platforms, subsequently increasing revenue raised from advertisements. Although for a long time, these services were wrongly considered "free", there is today a larger recognition of the market characteristics of these platforms, and even some movements towards compensating individuals more fairly for their contribution, such as data unions [13, 16, 18, 17], trusts [12, 15, 14], and the data dividend [4].

Black Markets for Data. Personal data such as passwords, and other authentication information, is often sold in the dark web. Although we do not focus on these black markets, it is important to recognize their existence. A prior qualitative analysis explained how data thieves maximize the number of fresh data copies sold by strategically reducing prices on data [38].

2.2 Data Markets in the Literature

In this section, we present work in economics and data management that engage with data's main characteristics: infinite replicability at low cost.

Digital Goods Markets. Goldberg et al. [37, 36, 35] studied truthful auction based mechanisms for revenue-maximization in the case of digital goods, which are infinitely replicable at zero cost. These studies observe that buyers' utilities are the only factor in maximizing revenue under infinite replicability. [37, 36] devised truthful competitive single and multiple-price auctions equivalent to setting the optimal fixed price on each item. [35] proved a tradeoff between truthfulness, revenue, and envy-freeness, and advocated randomized auctions as a mechanism to elicit almost-truthful bids from buyers while still ensuring revenue maximization.

Data Markets for Machine Learning (ML). The increased importance of ML techniques in the real-world has motivated development of ML-specific data markets. These works [20, 34, 39] consider the dataset exchange problem where the buyer's utility is the accuracy of the trained ML model. These studies assume sellers are cooperating and focus on fair revenue allocation via efficient computation of Shapley value. [20, 31] focus on disincentivizing strategic behavior based on infinite replicability. [20] combats the strategy of replicating the same feature to maximize revenue by proposing a version of the Shapley value that is robust to replication. [31] provides a mechanism to update prices temporally to combat the strategy of underbidding on the same data over time to purchase it at a reduced price.

Data Markets in Data Management. There has been a lot of work in the data management community [25, 24, 40, 42] to develop quoted price data markets that are designed for selling views of relational databases in a way that ensures no arbitrage[1]. Many recent proposals have studied data discovery and integration tasks [33], and pricing [30, 27, 43] for ML-related tasks. [43] com-

[1]No-arbitrage: it is cheapest to purchase a view directly instead of an equivalent combination of views

bines the Shapley value, arbitrage-free pricing, and differential privacy into a revenue-maximizing end-to-end marketplace for ML models. [30] propose completed federated Shapley value, an extension of Shapley value that allows two data owners with the same local data to receive the same valuation. A broader survey of analytical techniques for data pricing can be found in [48].

2.3 Online Data Marketplace Platforms

In its most basic form, data marketplaces offer an online list of datasets for sale and search functionality. Sellers contribute datasets and buyers can find them and purchase them. Datasets are often traded by money, but sometimes datasets are offered for free and the marketplace acts as a discovery platform. When sold for money, the pricing strategy varies from subscription-based pricing for cases where data freshness matters, to one-time transactions for a fixed price.

Datasets in these platforms rarely contain raw data about individuals, but commonly include aggregates ranging from socio-economic factors, to demographics and more [23]. Finally, most datasets are not sold exclusively in these platforms, and can instead be purchased directly by the sellers, that include many data vendors such as the data brokers we mentioned above. The most popular data marketplaces that enable easy access to datasets include Snowflake Data Marketplace [51], Narrative's Data Streaming Platform [9], Streamr Marketplace [11], AWS Data Exchange [22], and Dawex Global Data Marketplace [28]. Each of these platforms are popular among individuals from different industry sectors. Table 1 presents a summary of the different kinds of datasets traded in these marketplaces. Dawex and Narrative do not show complete lists of datasets, so we use the number of datasets displayed for a buyer account.

For our study we focus on two platforms, AWS, and Snowflake. These two offer the most transparency and that permits us to reach out to sellers and buyers to conduct structured interviews.

3. DATA TRANSACTION LIFECYCLE

In this section, we present the lifecycle of a data transaction from a sellers' and buyers' perspective. We construct the lifecycle (see Fig. 1) based on the studied platforms (i.e., AWS Marketplace and Snowflake) and informed by the interviews.

3.1 The Seller Pipeline

Identify Data To Be Sold. Sellers list datasets for sale that they believe will have a demand. Indeed, their competitive advantage is often that they possess market knowledge. When listing datasets, sellers must take into consideration potential legal constraints. For example, California and Vermont have increased restrictions on data trading through their data broker laws, which define the term "data broker" and require that data brokers register with the state, and disclose any security breaches [46, 47]. These laws apply to sellers of consumer data.

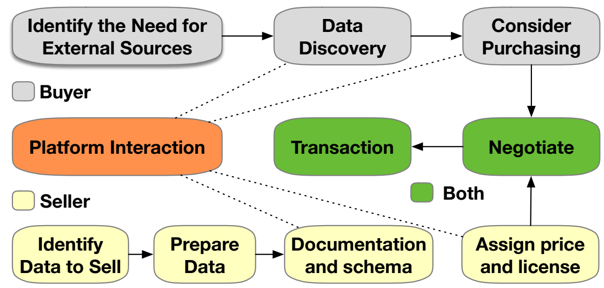

Figure 1: Data Trade Transaction Lifecycle. Dotted lines denote buyers' and sellers' reliance on the marketplace platform for respective components.

Prepare data. Before publishing the datasets, sellers generally perform data cleaning and processing to prepare it for potential consumers. Although preparing data without a concrete objective or a customer is hard, certain transformations such as removing personally identifiable information, removing duplicate values and encoding categorical attributes are often useful, e.g., M/F to Male/Female, or 0/1 to non-existent/existent values.

Documentation and Schema. Including metadata that describes the dataset is a key component of the seller's pipeline. According to the Arrows' information paradox [21], buyers do not buy datasets that they do not know help them, while sellers do not release data without a payment. This is one of the main challenges that hinders frictionless transactions among different entities. Including metadata that explains the dataset and how to use it is a way of signaling the value of the dataset to potential buyers. Most marketplaces require certain fields to be documented and allow sellers to include additional details that maybe helpful for buyers.

Assign price and license. Sellers must choose a license for the published data. Most marketplaces offer a predefined list, and permit sellers to include bespoke licenses when necessary. Finally, sellers must choose a price for the dataset based on their understanding of its value and popularity. However, current marketplaces do not provide tools to help sellers price datasets, e.g., demand in the form of customer search logs indicating intent.

3.2 The Buyer Pipeline

Identify the need for external data sources. Buyers participate in markets when they have a data need they cannot satisfy with the data they already own. The degree to which buyers can explain their data need differs among use cases. In some cases it may be very spe-

Industry	Ag/Envi- -ronment	Audience Marketing	Automotive	Corporate	Energy	Entertainment /Media	Finance	Health	Logistics	Manufacturing	Public Sector	Transportation /Travel	Telecomm- -unication
Common Features	pollution levels, precipitation	customer profiles, transactions	dealership performance	employee count, corporate linkage	oil levels, resource depletion rates	film reviews, news archives	credit scores, investment /trading	disease registry, virus tracking	imports /exports, supply chains	machine performance	tax rates, education performance	accident reports, airline traffic	cell tower data, signal strength
Dawex	> 90*	> 149*	> 339*	> 130*	> 207*	> 303*	> 581*	> 234*	> 201*	×	> 139*	> 196*	×
Narrative	×	> 24*	> 27*	> 1*	×	> 8*	> 50*	> 1*	> 1*	×	×	> 1*	×
AWS	×	1386	88	×	×	234	1362	379	×	85	638	×	237
Streamr	10	8	×	×	4	3	21	4	2	×	×	9	5
Snowflake	18	238	×	129	51	29	326	66	37	×	12	51	×

Table 1: Summary of data products and quantity available on data marketplaces as of May 2022.*estimates, as all datasets are not shown.

cific and concrete, such as when traders request tick data from the stock market. In contrast, it may be vague and harder to define, such as when a company wants to expand to new markets.

Data Discovery. Once in the marketplace, buyers get access to a simple, keyword-search-based search interface as well as to faceted search functionality that permits category-based searching such as 'file type', 'vendor', 'sector', etc.

Investigate potential purchase. Buyers can consult the metadata (documentation and schema) associated with datasets, but cannot directly observe data without payment. Samples of data may be included, but are not guaranteed. Usually, a buyer can also contact the seller to discuss their requirements.

3.3 The Transaction

Before a transaction takes place, buyers will often contact sellers to understand pricing, and clarify any characteristics of the dataset that are not available in the listing. Seller contact information is often included with each dataset listing and includes traditional methods such as phone numbers and email. We refer to this stage as negotiation, and include it in both the seller and buyer pipeline. Given the inconvenience of this stage in terms of time and effort, we performed user studies and interviews to understand seller and buyer engagement in data transactions and online marketplace use.

4. STUDYING DATA MARKETS

In this section, we present results of a survey of 30 sellers and 40 buyers, and interviews of 7 sellers and 16 buyers who actively participate in marketplaces (AWS and Snowflake). We requested interviews with all survey participants, some of whom are included in our interviews. We seek to understand whether current marketplaces fulfill their promise to distribute the value of data to wider audiences, and the challenges they face.

4.1 Study Methodology

We collected data from sellers and buyers via a survey and structured interviews. The interviews let us follow-up with participants and complement the information we collected in the survey.

Recruiting participants. We obtained consent from all participants and did not collect any personal information. To recruit sellers we obtained their publicly offered contact information from vendor websites in online marketplaces and from the list of brokers registered in California according to CCPA [46]. Recruiting buyers was difficult because marketplaces do not offer transaction information so we did not have a direct way of knowing who is buying data. We again relied on the CCPA broker registry and a spreadsheet provided as part of previous work [29] to find potential buyers, meaning the buyers are brokers. Both buyers and sellers come from a wide range of industries such as automotive, finance, etc. We had a low response rate, partially due to previous studies portraying brokers in a negative light [26].

4.2 Study Results

We present the results by presenting quantiles where the first or second quantile indicated either (i) agreement on the likert scale (3 or higher) (ii) 20% or more, depending on the question format.

4.2.1 Buyer Results

BQ1: For what purpose do you buy data? Participants were given the following list of categories to choose from: customer acquisition, efficiency of operations, employee satisfaction, reducing unnecessary expenses, resale, etc. More than 50% indicated that they purchased data mostly for customer acquisition, and around 25% indicated resale as the reason for the purchase. These results highlight the strong presence of brokers in data marketplaces, and hence, their presence in our sample.

BQ2: How do you decide if a price is worth it? How do you value a dataset? We provided buyers with the following characteristics of data and asked about their importance on a likert scale from 1-5 to help with understanding value. i) *metadata characteristics*: high-level dataset statistics included in the listing like titles, descriptions, data samples, identifiable vendor, open license, subscription plan, usage examples; ii) *hidden characteristics*: characteristics that only become apparent after negotiation rounds with sellers (such as quality e.g.

number of missing values, price, size, update frequency, geographical source, rareness).

More than 75% of buyers indicated that data title and usage description are helpful (5 on the likert scale). More than 75% of buyers indicated that they purchase datasets only when a seller is marked as an *'identifiable vendor'* (≥ 4 on likert scale). Choosing an identifiable vendor helps reduce scams. This highlights the importance of vendor reputation and personal relationships in marketplaces. Suggested usage of a dataset was not considered an important characteristic to estimate its value.

Regarding dataset characteristics, buyers pay most attention to the price ($> 75\%$ participants rated it to be the most important characteristic), followed by data quality (around 50% of the participants rated it 4 or above), of which they mostly refer to the existence of missing values. They do not weight as much other characteristics such as dataset size, update frequency, and geographical source of data.

BQ3: How much work do you do on the dataset after acquiring it?

Participants were provided a list of actions and asked to provide the percentage of time spent in the following aspects of a transaction: negotiating license, receiving files, familiarizing themselves with data, capitalizing on insights, sending the data to others, etc. More than 75% of the participants indicated that they spend the most of their time familiarizing themselves with the data, once it is acquired.

BQ4: What's the value of data marketplaces for buyers? What do they lack?

Participants mentioned that most time is spent in negotiations with the sellers, followed by figuring out license negotiations and data transfers. They indicated the main value of marketplaces is to obtain access to a centralized list of datasets. We asked survey participants which data marketplace features among keyword search, vendor filter, industry filter, file type filter, descriptions, contact information and metadata they consider necessary. More than 75% of participants find metadata most valuable, followed by concrete contact information and search filters. The latter emphasizes the expectation that a one-on-one negotiation will be necessary before a transaction takes place.

4.2.2 Seller Results

SQ1: Why do you sell data? Do you experience long-term benefits to sharing your data beyond the initial transaction?

Every participant had differing opinions on the impact that making data more available would have on the economy. Yet, all of them agreed that monetary gains is the main reason to sell data – this is the main business model of data brokers.

SQ2: How much effort goes into sharing a dataset for sale? Preparing, removing PII, etc.

Participants were provided a list of actions and asked to provide the percentage of time spent. More than 75% of participants indicated they generally spend the maximum amount of time on data preparation tasks like formatting, filling missing values, layout changes (e.g., splitting a column into two). Sellers spent less time on documentation because they do not know what documentation is useful for buyers (learned from interviews).

SQ3: How do you choose a price for the dataset? Participants were provided various criteria for pricing datasets to choose from: collection date, size, quality, market demand, documentation, likelihood of similarity to other datasets on the market, prices of other data sold before, resources spent for collection. Most participants indicated that "demand" is the primary criteria that drives prices. Data brokers are in the marketplace because they have already identified an opportunity to sell data. Other factors that influence prices include data collection costs, data errors, and dataset uniqueness.

SQ4: What is the most time-consuming, costly part of the process of selling data?

Participants were provided a list of actions and asked to provide the percentage of time spent: negotiating price, negotiating license, finalizing sale, transferring data, technical issues, other. More than 50% of participants mentioned they spend at least a quarter of the effort negotiating a price for the transaction, even though there was high variance among respondents. Better known datasets sold regularly require less negotiation, while newer datasets take longer. Sellers mentioned that price is one of the main reasons why transactions fall apart.

SQ5: What's the value of marketplaces for sellers? Participants were provided options such as: choosing among open and configurable licenses, choosing between pricing options, assistance in determining how much data is worth, sensitive data identification, publishing data easily. Publishing data received one of the highest scores– 27 of 30 participants rated it 4 or 5 out of 5, reflecting that sellers leverage marketplaces for publicity and to gain exposure. However, more than 75% of participants indicated that having some assistance to select a price for their data would be helpful. We did not hear any proposal to achieve that, and it remains one of the most difficult aspects of a transaction.

4.2.3 Result Synthesis

Today's online data marketplaces act as a storefront where sellers have an opportunity to list datasets and buyers a central place to search for opportunities. The main value of these marketplaces is in reducing search costs. The transaction cost is high, as both buyers and sellers emphasize they must negotiate sales one-on-one, and that this process remains mostly manual (BQ2, BQ4, SQ4). Buyers find metadata describing the datasets and contact information most useful, while sellers would benefit from strategies to price their datasets (BQ2–BQ4, SQ2). Their current strategy takes into consideration demand, as sellers consist mostly of organizations that have already identified an opportunity to trade data (SQ3). Additionally, devising tools to identify useful metadata to document and publish datasets (SQ2) and strategize pricing based on its demand and popularity (SQ4) would help sellers.

5. NEXT STEPS FOR DATA MARKETS

One of the key requirements of a well-functioning data market is that buyers and sellers engage in frictionless transactions. This permits the spread of the value of data to wider audiences effectively and efficiently. One mechanism to ensure this is with low transaction costs, which is possible if the markets are *thick* [49], meaning there are many transactions and consequently search costs are low. Increased transactions mean that more people benefit from the available datasets. It also simplifies the pricing aspect of data, as prices tend to settle to clear the market, i.e., to supply the existing demand. As we found in BQ2, BQ4, and SQ4, one-on-one negotiation of each sale is the main bottleneck to allow smooth transactions. In fact, negotiations take more effort than other tasks (BQ2, SQ4). Below are some measures markets could take to improve transactions and to converge towards better (frictionless) data markets.

Reducing search effort for buyers. Current systems require buyers to discuss their needs with sellers to understand the value of available data, which is time consuming (BQ2-BQ4). If buyers instead find it effortless to find relevant data, they will have more incentive to search for that data in the market. This requires datasets to be accompanied with metadata and development of sophisticated tools to allow buyers to estimate usefulness of datasets without the overhead of an interaction.

Reduce publishing cost for sellers. Current data markets do not offer sellers much help to document and price datasets (SQ2-SQ4). Designing tools to reduce the effort needed to populate metadata and publish valuable data would reduce friction. If sellers find they profit from sharing datasets, they will be incentivized to share.

Below, we provide our perspective on opportunities to exploit existing technological advances, create new innovative solutions to data management problems and engineer better data markets.

Data as a service. A fundamental challenge of data markets is that buyers do not pay without knowing the value they get from a dataset, and sellers will not release the dataset before receiving a payment. This is because, unlike with rival goods (that cannot be freely replicated), once buyers see the data, they can copy it and lose any incentive to pay sellers. Although legal mechanisms could be put in place, these are precisely the kinds of manual mechanisms that motivate the one-on-one negotiations precluding frictionless transactions. Instead, we argue that data's value is often *instrumental*: data is valuable as a means to an end, e.g., valuable to complete the buyer's task. This suggests a path towards addressing this issue: offer data as a service, by requesting the service from buyers and automatically finding datasets that satisfy the task. Buyers can price how valuable they find the task solution, and the market can match supply and demand. We are working on a new platform that would support these *data-blind interfaces*, as explained in previous work [33, 32]. We believe these mechanisms are a way of achieving low transaction cost that would incentivize buyers to participate, hence increasing market thickness.

Curation tools. One component of publishing cost is ensuring privacy: a seller will only share their dataset for sale if the advantages of selling the dataset outweigh the risk of leaking sensitive and private information. Services to prepare datasets that protect sellers by helping manage privacy constraints as well as populating metadata will enrich markets, increase seller trust and incentivize more sellers to share their data.

Market Robustness. We end with a note on *robustness* [49] of markets to strategic attacks from adversarial self-interested individuals. Recent studies reveal attacks that could affect the normal course of transactions [20, 31] and theoretical mechanisms for a robust market design. A non-robust market will quickly disincentivize honest participants, reducing transactions and breaking the market. Therefore, robustness is crucial in designing marketplaces, assuring sellers that they received the best price, and buyers that they did not overpay.

6. CONCLUSION

In this paper, we studied the most popular data marketplace platforms. We analyzed the data sold in the platforms, and then interviewed both buyers and sellers to understand their experience. We gained insight into both the lifecycle of data trading, which we presented here, and the major roadblocks to frictionless data trading. We presented the results of the interviews conducted and then connected them to what we believe are the major challenges to allowing the effective spread of data value in data marketplaces. We offered a few avenues of future research.

7. REFERENCES

[1] Amazon https://amazon.com.

[2] Bloomberg https://www.bloomberg.com/.

[3] California consumer privacy act (ccpa) https://oag.ca.gov/privacy/ccpa.

[4] Data dividend project https://www.datadividendproject.com/.

[5] Facebook {https://facebook.com}.

[6] Gmail https://mail.google.com/mail/.

[7] Instagram {https://instagram.com}.

[8] Morningstar inc https://www.morningstar.com/.

[9] Narrative. https://www.narrative.io/.

[10] Netflix https://netflix.com.

[11] Streamr. https://streamr.network/.

[12] "data trusts" could be the key to better ai https://hbr.org/2020/11/data-trusts-could-be-the-key-to-better-ai, 2020.

[13] Data unions research summary https://streamr.network/resources/data_unions_research_report.pdf, Feb 2020.

[14] The future of data trusts and the global race to dominate ai https://www.bennettinstitute.cam.ac.uk/blog/data-trusts1/, 2021.

[15] How data trusts can protect privacy https://www.technologyreview.com/2021/02/24/1017801/data-trust-cybersecurity-big-tech-privacy/, 2021.

[16] Swash: Reimagining data ownership https://swashapp.io/, Jan 2021.

[17] Tx project: Tracey https://tx.company/projects/tracey/, Jan 2021.

[18] Welcome to the diem project https://www.diem.com/en-us/, Jan 2021.

[19] Acxiom. About acxiom https://www.acxiom.com/about-us/, Jan 2021.

[20] A. Agarwal, M. Dahleh, and T. Sarkar. A marketplace for data: An algorithmic solution. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, EC '19, pages 701–726, New York, NY, USA, 2019. ACM.

[21] K. J. Arrow. *Economic Welfare and the Allocation of Resources for Invention*, pages 219–236. Macmillan Education UK, London, 1972.

[22] Aws data exchange. https://aws.amazon.com/data-exchange/.

[23] S. A. Azcoitia and N. Laoutaris. A survey of data marketplaces and their business models, 2022.

[24] S. Chawla, S. Deep, P. Koutrisw, and Y. Teng. Revenue maximization for query pricing. *Proc. VLDB Endow.*, 13(1):1–14, Sept. 2019.

[25] L. Chen, P. Koutris, and A. Kumar. Towards model-based pricing for machine learning in a data marketplace. In *Proceedings of the 2019 International Conference on Management of Data*, SIGMOD '19, pages 1535–1552, New York, NY, USA, 2019. ACM.

[26] F. T. Commission. Data brokers: A call for transparency and accountability: A report of the federal trade commission (may 2014) https://www.ftc.gov/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014, May 2014.

[27] Z. Cong, X. Luo, J. Pei, F. Zhu, and Y. Zhang. Data pricing in machine learning pipelines. *CoRR*, abs/2108.07915, 2021.

[28] Dawex: Sell, buy and share data. https://www.dawex.com/en/.

[29] M. Elsaify and S. Hasan. Some data on the market for data. *IO: Empirical Studies of Firms & Markets eJournal*, 2020.

[30] Z. Fan, H. Fang, Z. Zhou, J. Pei, M. P. Friedlander, C. Liu, and Y. Zhang. Improving fairness for data valuation in federated learning. *CoRR*, abs/2109.09046, 2021.

[31] R. C. Fernandez. Protecting data markets from strategic buyers. *SIGMOD*, 2022.

[32] R. C. Fernandez, K. Chard, B. Blaiszik, S. Krishnan, A. Elmore, Z. Obermeyer, J. Risley, S. Mullainathan, M. Franklin, and I. Foster. The data station: Combining data, compute, and market forces, 2020.

[33] R. C. Fernandez, P. Subramaniam, and M. J. Franklin. Data market platforms: Trading data assets to solve data problems. *PVLDB*, 13(12):1933–1947, jul 2020.

[34] A. Ghorbani and J. Zou. Data shapley: Equitable valuation of data for machine learning, 2019.

[35] A. Goldberg and J. Hartline. Envy-free auctions for digital goods. 09 2003.

[36] A. Goldberg, Jason, H. May, A. V. Goldberg, and J. D. Hartline. Competitive auctions for multiple digital goods. In *In Proc. 9th European Symposium on Algorithms*, pages 416–427, 2001.

[37] A. V. Goldberg, J. D. Hartline, and A. Wright. Competitive auctions and digital goods. In *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '01, page 735–744, USA, 2001. Society for Industrial and Applied Mathematics.

[38] T. J. Holt and E. Lampke. Exploring stolen data markets online: products and market forces. *Criminal Justice Studies*, 23(1):33–50, 2010.

[39] R. Jia, D. Dao, B. Wang, F. A. Hubis, N. M. Gurel, B. Li, C. Zhang, C. Spanos, and D. Song.

Efficient task-specific data valuation for nearest neighbor algorithms. *Proc. VLDB Endow.*, 12(11):1610–1623, July 2019.

[40] P. Koutris, P. Upadhyaya, M. Balazinska, B. Howe, and D. Suciu. Query-based data pricing. *J. ACM*, 62(5):43:1–43:44, Nov. 2015.

[41] P. Koutroumpis, A. Leiponen, and L. D. W. Thomas. Markets for data. *Industrial and Corporate Change*, 29(3):645–660, 03 2020.

[42] C. Li, D. Y. Li, G. Miklau, and D. Suciu. A theory of pricing private data. *ACM Trans. Database Syst.*, 39(4):34:1–34:28, Dec. 2014.

[43] J. Liu, J. Lou, J. Liu, L. Xiong, J. Pei, and J. Sun. Dealer: An end-to-end model marketplace with differential privacy. *Proc. VLDB Endow.*, 14(6):957–969, feb 2021.

[44] A. Muschalle, F. Stahl, A. Löser, and G. Vossen. Pricing approaches for data markets. In *International workshop on business intelligence for the real-time enterprise*, pages 129–144. Springer, 2012.

[45] Nielsen. Nielsen marketing cloud: Real science. real results., Jan 2021.

[46] S. of California Department of Justice. Data broker registry, `https://oag.ca.gov/data-brokers`. *State of California Department of Justice*, Jan 2021.

[47] V. S. of State. Data brokers, Jan 2021.

[48] J. Pei. A survey on data pricing: from economics to data science. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–1, 2020.

[49] A. E. Roth. What have we learned from market design? *The Economic Journal*, 118(527):285–310, 2008.

[50] F. Schomm, F. Stahl, and G. Vossen. Marketplaces for data: an initial survey. *ACM SIGMOD Record*, 42(1):15–26, 2013.

[51] Snowflake data marketplace. `https://www.snowflake.com/data-marketplace/`.