Databases and Information Systems VIII
H.-M. Haav et al. (Eds.)

doi:10.3233/978-1-61499-458-9-145

Data Marketplaces: An Emerging Species

Florian STAHL [a,1], Fabian SCHOMM [a], and Gottfried VOSSEN [a,b]
[a] *European Research Center for Information Systems (ERCIS)*
University of Muenster, Germany; firstname.lastname@*uni-muenster.de*
[b] *The University of Waikato Management School, New Zealand, vossen@waikato.ac.nz*

Abstract. Trading data as a commodity has become increasingly popular. To obtain a better understanding of the emerging area of data marketplaces, we have conducted two surveys to systematically gather and evaluate their characteristics. This paper essentially continues and enhances a survey we conducted in 2012; it describes our findings from a second round done in 2013. Our study shows that the market is vivid with numerous exits and changes in its core business. We try to identify trends in this young field and explain them. Most notably, there is a definite trend towards high quality data.

Keywords. Data Marketplaces, Data Marketplaces Survey, Data Trading

Introduction

With more and more data being created, analyzed, and stored across virtually all business areas, it becomes increasingly difficult to have the appropriate data present at the right time. Furthermore, data quality is an ongoing issue in many technology-driven applications. Due to the abundant supply of data nowadays, finding the data that is most apt for someone's needs is often very challenging, especially when facing the need for regular updates and a sufficient level of data quality. This challenge has led to the emergence of specialized companies that promote the proliferation of data as a commodity by offering services that we call *data marketplaces* [1,2]. As is common in newly established fields, this market of data providers is crowded with a vast numbers of competitors, varying in size, focus, and other dimensions. This leads to uncertainty in data consumers, who are faced with the task of finding and selecting the right data marketplace for their needs. The purpose of our work is to shed light on this matter.

In our earlier work, a basic framework for the categorization of data vendors and data marketplaces has been introduced, followed by a survey of the current market situation in 2012 [2]. In this paper, this framework is extended to incorporate new research insights, and the results of the reiterated survey performed in 2013[2] are presented. This paves the way for a comparative study that aims at identifying trends and shifts by contrasting the results from with 2012 with those from 2013. As it turns out, only one year was enough time to find meaningful results in this regard. Additionally, a co-occurrence analysis has

[1] Corresponding author
[2] The precise time frame of the survey was from July to August 2013.

been added to this study that aims at revealing common combinations of characteristics among the survey participants.

This paper is a revised version of [3][3]. Its structure is as follows: First, the approach and differences between both surveys is described in Section 1. Then, the findings of the recent survey from 2013 are presented in Section 2. Section 3 explains the most relevant related work in this field. Finally, the conclusion as well as our prediction of trends regarding data marketplaces is given Section 4.

1. Methodology and Approach

In this section, the methodological similarities and dissimilarities with our previous survey will be outlined, followed by short remarks about the data acquisition process. After that, the basics of the comparative study as well as the co-occurrence analysis are described.

1.1. Similarities and Dissimilarities with the Previous Survey

In order to make this survey vastly comparable to our previous work [2], only slight modifications have been made to the methodology. In particular, our definition of *data marketplace* and *data vendor* remain unchanged. In short, the focus is laid on companies offering either a platform that allows users to buy and sell (or just offer) data (e. g., datamarket.com), providing raw data in any form (e. g., data.gov), and on companies offering data enrichment tools (e. g., attensity.com). A further selection criterion for the companies surveyed has been that they offer their products and services over the Internet. For a more comprehensive definition, interested readers are referred to our previous work [1,2].

The general limitations of this type of survey have not changed. They are as follows: (1) The data collected comes purely from the Web sites of the respective vendors. While we are aware that vendors try to present themselves in the best way, which might lead to a bias in the findings, a more thorough data collection step (i. e. actual testing of the offered services) is prohibitive in our case due to lack of the resources needed. In this sense, width has been prioritized over depth. (2) As it was not possible to find information about every dimension for every vendor, the completeness of the survey data is not perfect. However, we refrained from guessing missing values, and instead treated these fields as `NULL` values. This could lead to minimally skewed results, but it is our belief that this approach yields the best overall results. (3) Our survey does not claim to be exhaustive, as that would be close to impossible due to the sheer number of different actors and vendors in this field. Nevertheless, we have tried our best to survey the most important vendors as well as representative niche vendors. We believe that this gives a pretty accurate view of the overall market situation[4]. Companies that feel they have been left out although they should be included here are encouraged to contact and inform us.

[3] An extended abstract appeared in Proc. 11th Int. Baltic Conf. on DB & IS 2014, 135–146.

[4] The list of companies surveyed can be found at `http://dbis-group.uni-muenster.de/y/survey2list`, and we are happy to provide the full data of the survey upon request.

1.2. Data Acquisition and Approach

Given that this is essentially a continuation of a previous study, all previously surveyed vendors have been revisited (cf. [2]). The actual data acquisition was performed through means of an online investigation. In order to speed up this process, the number of surveyors was doubled from last year. Similar to last year's survey, new candidates for examination have been search by means of a broad keyword-based search. Additionally, suggestions and feedback from peers with whom we discussed our previous work have been considered.

For continuity reasons, the analysis follows the same twelve dimensions that have been established in the previous survey, which are divided into objective and subjective dimensions. During the investigation phase, however, it became apparent that these dimensions do not cover every interesting aspect and thus, need to be extended. Therefore, two new dimensions have been added, namely *Pre-Purchase Information* (subjective) and *Pre-Purchase Testability* (objective).

As in the previous survey, all values are strictly Boolean. An offering either fulfills the criteria for a certain dimension category or it does not. However, categories are not mutually exclusive in most cases, e. g., one offering can provide multiple ways of data access. Dimensions that are mutually exclusive will be pointed out in their respective description in Section 2.

1.3. Method of Comparative Study

The market for data vendors and data marketplaces is an emerging and dynamic market. Thus, some offerers leave the market and new providers appear. We observed that out of all companies that were surveyed last year, three companies went out of business and one changed their core offering so much that it no longer fits our definition of a data marketplace.

In order to properly account for the changes in this market, we also looked at new companies to include. It turned out that 5 companies, which had not been part of the previous survey, fit nicely into our selection criteria as explained in Section 1.1. This results in 3 distinct groups of companies surveyed which we gave the following names:

- *Leavers*: Companies that took part in the previous survey, but no longer exist or have changed business now — 4 companies.
- *Returners*: Companies that are continuously part of the survey — 42 companies.
- *Freshmen*: Companies new to the survey — 5 companies.

Returners and *Freshmen* together build the basis for the 2013 data which will be — similar to last year — described in the findings section (Section 2). In the same context, the developments of the market situation are highlighted. To this end, the data gathered in 2012 for the Returners is compared with the 2013 survey data for this group. By doing so, an initial picture of how the market is developing can be drawn. It is obviously still too early to take these two surveys as a basis for serious predictions, but we are confident that our results form a solid foundation for further research in the future.

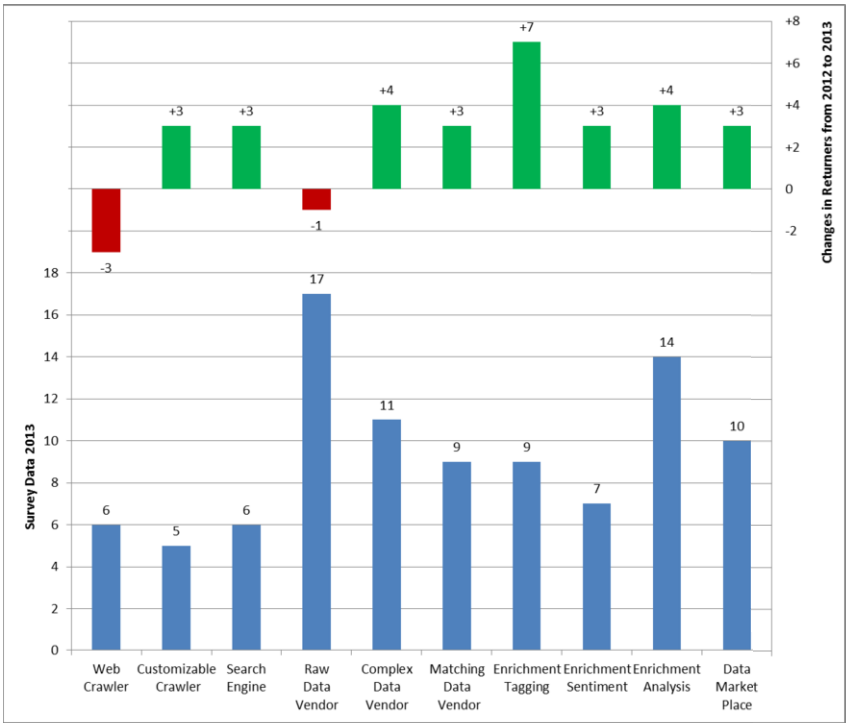

Figure 1. Number of Vendors for each Type

1.4. Co-Occurrence Analysis

When analyzing subjects with multiple attributes (in this case, categories), the question arises whether there are meaningful relations between those attributes. In order to answer this question, we performed an analysis of the co-occurrence of certain categories, i. e. what are the most common combinations of categories across dimensions. The Jaccard index, which calculates similarity of sets, is well suited for this task. Given two sets A and B, the well-known Jaccard index $J(A,B)$ is defined as $J(A,B) = \frac{|A \cap B|}{|A \cup B|}$. In our case, the sets are the categories of the individual dimensions, and their overlap is the number of data vendors that fulfill the criteria for both categories. To make sure that we present verifiable and meaningful results, we decided to focus on the objective dimensions and omit the subjective dimensions as well as the language dimensions from this analysis step. The results of the co-occurrence analysis are presented in Section 2.3.

2. Findings

The results for 2013 will be presented in a similar manner as was done in [2]. In contrast to the previous survey, no in-depth explanation of each category will be given except for the two new dimensions (cf. Section 1.2). The interested reader is referred to a glossary of important terms in [3]. Also, the presentation here is focused on the development over the course of one year instead of only showing the figures for 2013. To that end, all following

charts are split into two parts. The lower chart shows the current findings from 2013 in absolute numbers, i. e. total count of vendors fulfilling that category. The upper chart illustrates the changes within the group of returners[5]. This implies that the situation of 2013 depicts more than the survey of 2012 plus changes, as also new vendors have been included in this survey. More concretely, the lower bar chart is based on the two groups returners and freshmen, and thus, has a maximum value of 47. The upper chart on the other hand represents the change in returners, with a maximum value of 42.

2.1. Objective Dimensions

2.1.1. Type

The *Type* dimension classifies vendors regarding their core offering. As shown in Figure 1, this non-mutually exclusive category exhibits a growth in all categories. The strongest increase is in *Enrichment Tagging*. Generally, it can be seen that products that offer enhanced data are increasingly common, while the types of services that decreased slightly are offering unprocessed information, i. e. *raw data* or *non-customized crawling*. These numbers might indicate a shift from crude data towards refined data.

2.1.2. Time Frame

Time Frame describes the temporal context of the data. It can be broadly distinguished between data that is valid and relevant for a long period of time (*static/factual*) and data which is only valuable shortly after its creation (*up to date*).

The most interesting result in this dimension is that the percentage of vendors offering both static and up-to-date data increased from less than 20% in 2012 to approximately 45% (21 vendors) in 2013 (cf. Figure 2, the lower part of the bar shows the number of vendors fulfilling both categories). Also, the gap between both shrank from 9 to 4 vendors in the overall sets. Additionally, this trend is evident by the stronger increase in up-to-date information within the returners group.

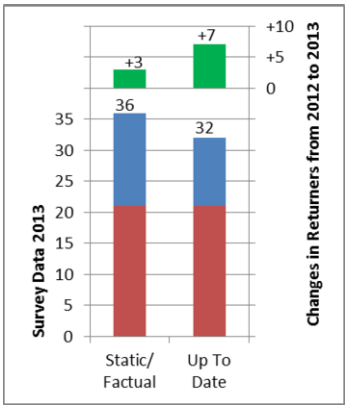

Figure 2. Number of Vendors for Time Frame

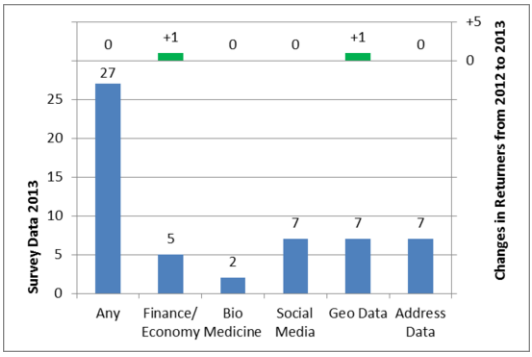

Figure 3. Number of Vendors for each Domain

[5]The two newly added dimensions do not have these changes because there is no previous data to compare to.

2.1.3. Domain

The *domain* describes the business area to which the data is related. The domain *any* describes vendors who do not specialize in any particular domain (for instance on a data marketplace, where all kinds of different data is available). As in 2012, vendors falling into the *any* category did not count towards the other, more explicit domains. In contrast, other domains were not mutually exclusive, i. e., a vendor may serve more than one domain, which is commonly the case with *Geo Data* and *Address Data*. Figure 3 shows that the situation did not change much at all compared with the previous study. This underpins the impression that changing of the domain is not a short term undertaking.

2.1.4. Data Origin

Data origin describes the source from which the data comes.

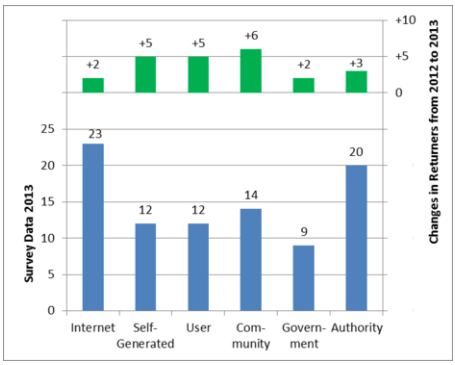

Figure 4. Data Origin Distribution

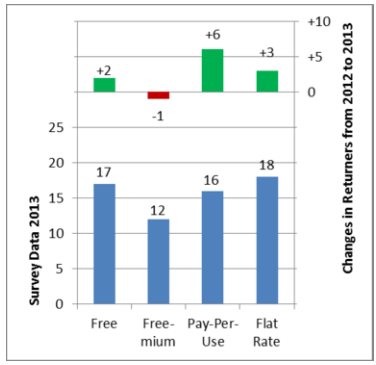

Figure 5. Number of Vendors for each Pricing Model

In the 2013 survey, *Internet* and *Authority* stay the most popular sources of data with 23 and 20 vendors, respectively. Despite the fact that the main advantage of these offers is that the data is usually of high correctness, completeness, and credibility, we observed an increase of more than 80% in origins *self-generated*, *user*, and *community* within the group of returners (cf. Figure 4). Regarding *user*, which consists mainly of enhancement services, this may suggest an increased need in adding value to the data a company has already at hand. Also, the raise in *self-generated* and *community* may suggest that there is a need for data that cannot be generated or sourced by other means. Overall though, rising numbers have been observed for all categories, which indicates that the market is still growing rapidly. It remains to be seen how this development shakes out.

2.1.5. Pricing Model

The four *pricing models* described last year are still those most commonly used. As can be seen in Figure 5, the distribution among them is fairly even, with the exception that *freemium* is under-represented. Interestingly, *freemium* has lost importance while at the same time pay-per-use increased strongly, both within the returners and regarding the overall set. This development indicates that the willingness to pay for data is increasing.

2.1.6. Data Access

Data access describes, how end-users receive the data from vendors. In this regard, *APIs* remain the most widely offered means of accessing data, although the count did not increase. Quite interestingly, the proprietary access through *specialized software* has the strongest relative increase with about 60% for the returners group, nevertheless staying the least frequently offered means of access in absolute numbers. Figure 6 presents the details. A new observation is that, unlike in the last survey, some vendors could be detected that offer all types of data access. Even though their number is still relatively small (5 or ∼11%), this might be the beginning of a trend towards ubiquitous data access, which lets the customer choose his preferred access method.

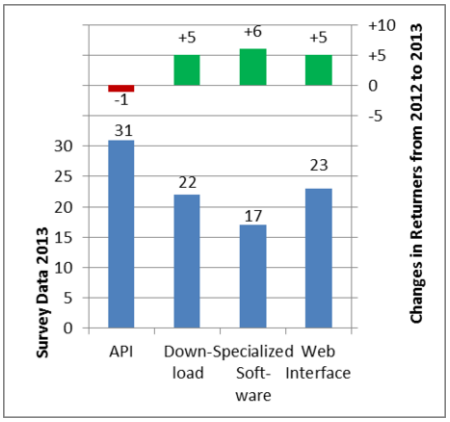

Figure 6. Data Access Distribution

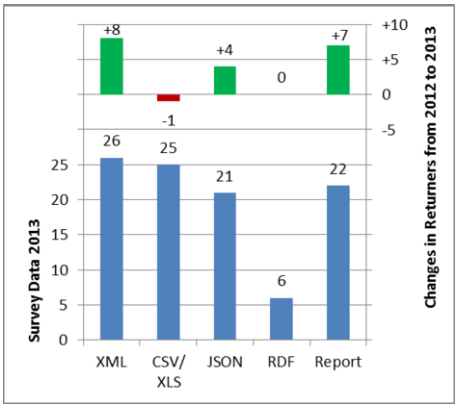

Figure 7. Number of Vendors per Data Output Category

2.1.7. Data Output

Figure 7 shows the format in which data can be obtained. As can be seen, *XML* superseded *CSV/XLS* as the most popular data output format. Together with the increase for *JSON*, the assumption can be made that Web standards are about to replace the more traditional exchange formats. Two vendors even offer all data output formats. The increase in pre-formatted reports is feeding the impression from previous sections that vendors try to individualize themselves as well as providing immediate benefits by simplify data access for managers and other non-technical personnel.

2.1.8. Language

As in the previous study, the language analysis distinguishes between the language of Web sites and the language of the data itself. In Figure 8, it can be seen that English is still the dominant Web site language with only minor increases in German and other languages with the returners, which is unsurprising. However, looking at the language of the data, it can be seen that English has only a little growth while German and other languages increase significantly. This suggests that there is a rising demand for national, non-English data.

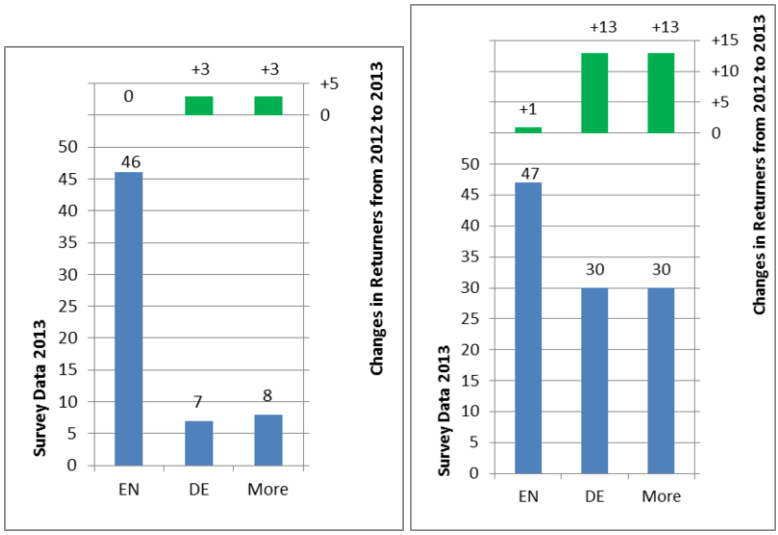

Figure 8. Language of Web Sites (left) and Data (right)

2.1.9. Target Audience

The dimension *Target audience* captures whether an offering focuses on business customers (B2B) or consumers (B2C)[6]. Figure 9 illustrates the numbers for both categories. The number of vendors offering services in both categories increased, in this case from 28% to 43% in the overall set. At the same time, it could be observed that more than twice as much offerings focus on business customers than on consumers. We conclude that data services currently are — and most likely will remain — a B2B-centric market.

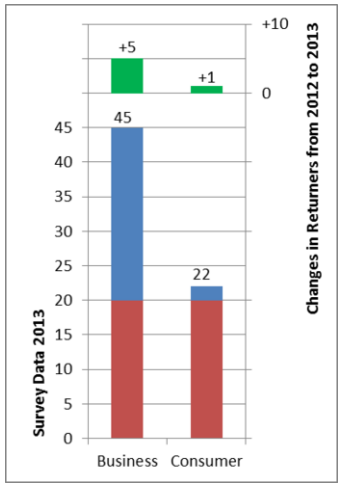

Figure 9. Number of Vendors by Target Audience

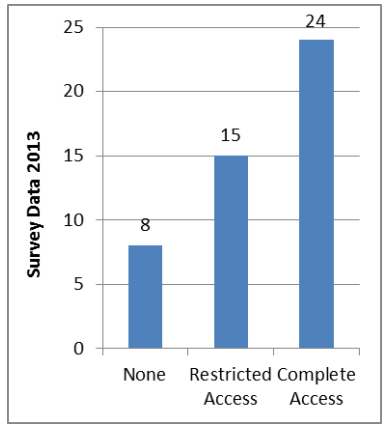

Figure 10. Number of Vendors by Pre-Purchase Testability

[6] *Consumers* was last year referred to as (end) customer. For clarity, we renamed this category to *Consumers*

2.1.10. Pre-Purchase Testability

This is one of the two new dimensions of this year's survey (depicted in Figure 10). It describes to what extent data offerings can be tested before an actual purchase is made. From our survey we derived the three intuitive categories:

- *None*: No access is given to the data before purchase, leaving the demanders to buy the proverbial pig in a poke.
- *Restricted Access*: In this category, pre-purchase access to the service is either limited by time (e. g., 30 days trial) or by API calls/data volume (e. g., first 100 calls / 100MB free, a typical implementation of freemium pricing models).
- *Complete Access*: Vendors in this category allow a complete access before purchasing

Assuming that most buyers are interested in as much information as possible before they purchase a service, it does not come as a surprise that more than 80% (39 vendors) of the sample vendors offer at least restricted access. Even though 17% (8 vendors) seems a rather low figure compared to that, it is an unexpectedly high total number given that these vendors expect their customers to rely on the vendor's promises and do a blind bargain.

2.2. Subjective Dimensions

2.2.1. Trustworthiness

For this dimension we assessed the trustworthiness of vendors based on the origin of their data as well as on how it is processed. As in 2012, this dimension is not meaningfully quantifiable and, thus, the results are subjectively biased. Also, we kept the method of allowing multiple entries for one vendor as one vendor can offer multiple services or data sets with varying credibility. As can be seen in Figure 11, there is no clear trend recognizable. While an increase on both ends (i. e., barely and highly trustworthy) can be observed, at the moment any interpretation here would be sheer speculation.

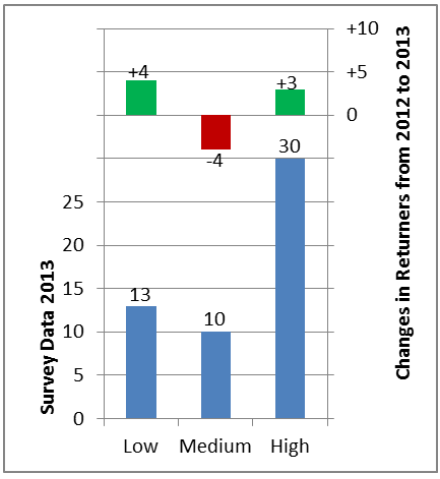

Figure 11. Trustworthiness Distribution

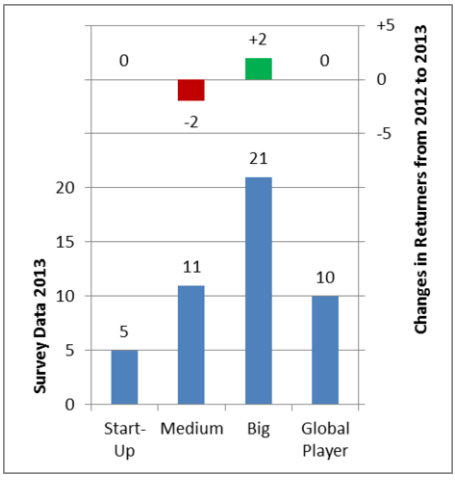

Figure 12. Number of Vendors by Size

2.2.2. *Size of Vendor*

Similar to the previous survey, we used a vendor's Web presentation as foundation for a classification regarding the size of a vendor, which is naturally mutually exclusive.

Figure 12 shows that the overall distribution has not changed much. Nevertheless, within the returners group an increase in size of vendors can be seen. Also, in the overall result for 2013 the relation of big to medium companies favors big companies more than it did in 2012, while the startup and global player remain about the same, which suggests that the market is growing and companies are developing.

2.2.3. *Maturity*

Similar to *Size of Vendor*, *Maturity* has not changed tremendously compared to 2012. Regarding the overall set, a minor increase in medium and high maturity can be observed. This maturing trend is also supported by the deltas for the returners. This is illustrated in Figure 13 and supports the suggestion made in the previous subsection that the market and companies are not only growing but also maturing, admittedly at a rather slow pace.

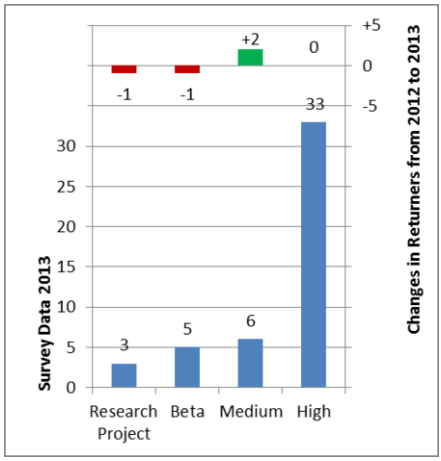

Figure 13. Maturity of Vendors

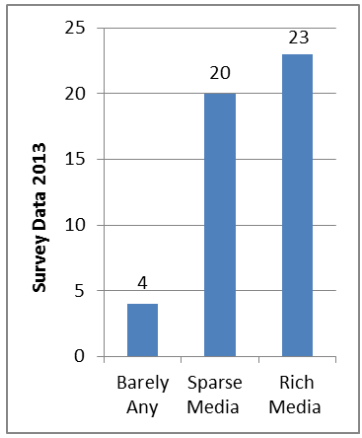

Figure 14. Number of Vendors by Pre-Purchase Information

2.2.4. *Pre-Purchase Information*

Pre-purchase information is the second new dimension. Unlike testability, which classifies hands-on experiences, this dimension evaluates how well and extensive a supplier provides information and details about his offering in advance to a purchase. This dimension is inherently subjective, as the same information might be differently interpreted by different people. For that reason we focused mainly on the extent — rather than the quality — of the information. Nevertheless, the values might be subjectively biased. The following three categories in this dimension have been observed:

- Barely Any (Information): The information given on the Web site is only textual and rather sparse. Potential customers are often asked to send their enquiries via e-mail for more details.

- Sparse Media (Information): This group compromises vendors offering more comprehensive textual and sparse media information such as short video demonstrations.
- Rich Media (Information): Besides extensive textual description of their services, these vendors supply a plenitude of media information such as screen casts.

Given that more information enables potential customers to evaluate a service more comprehensively to match it with their needs, it is little surprising that 4 vendors supply hardly any information beforehand. On the other hand, it is exemplary that nearly half (23 vendors) of the vendors surveyed in 2013 supply comprehensive media information to reduce their (potential) customer's uncertainty and facilitate a purchase decision. These numbers are visualized in Figure 14.

2.3. Co-Occurrence-Analysis Results

This section presents the results of the co-occurrence-analysis using the Jaccard similarity index as introduced in Section 1.4. Due to the simplicity of the formula, all calculations have been performed using standard spreadsheet software. Table 1 shows those combinations of attributes that have the highest Jaccard index. Note that due to the commutative property of the Jaccard index, the ordering of the dimension/category columns is arbitrary.

Table 1. Top 20 dimension/category combinations with highest Jaccard index.

#	Dimension: Category	Dimension: Category	Jaccard
1	Target Audience: Business	Time Frame: Static/ Factual	0,723
2	Target Audience: Business	Time Frame: Up To Date	0,711
3	Data Output: CSV/ XLS	Data Access: Download	0,679
4	Domain: Any	Time Frame: Static/ Factual	0,658
5	Target Audience: Business	Data Access: API	0,652
6	Data Output: Report	Data Access: Web Interface	0,607
7	Data Output: CSV/ XLS	Time Frame: Static/ Factual	0,605
8	Domain: Address Data	Category: Matching Data Vendor	0,600
9	Data Origin: Community	Type: Data Market Place	0,600
10	Data Access: API	Time Frame: Static/ Factual	0,595
11	Data Output: XML	Time Frame: Static/ Factual	0,590
12	Data Output: XML	Data Access: API	0,583
13	Target Audience: Business	Data Output: XML	0,578
14	Pre-Purchase Testab.: Complete Access	Pricing Model: Free	0,577
15	Data Access: API	Time Frame: Up To Date	0,575
16	Data Access: Download	Time Frame: Static/ Factual	0,568
17	Data Output: Report	Type: Enrichment - Analysis	0,565
18	Target Audience: Business	Domain: Any	0,565
19	Target Audience: Consumer	Pricing Model: Free	0,560
20	Domain: Social Media	Type: Enrichment - Sentiment	0,556

Dimensions *Target Audience* and *Time Frame* both have only two distinct categories and consequently have the highest total count of vendors. Combined with their non-mutual exclusiveness, the high Jaccard index does not surprise.

Examining the other results more closely, it becomes apparent that dimensions *Data Output* and *Data Access* have interesting combinations. These are *CSV/XLS* and *Download* (ranked 3rd), *Report* and *Web Interface* (ranked 6th), and *XML* and *API* (ranked 12th). This makes sense from a technological point of view: data that comes in tabular files is mostly processed further using desktop spreadsheet software and thus, most conveniently downloaded directly. Reports on the other hand are most often meant to be directly viewed by end-users, which is easiest done using a browser. Finally, formatting data using XML is most likely done with the intention to expose that data through an API, because it is easily machine-readable. So even while one might have guessed that there is a high likelihood for these combinations, our numbers verify this.

Dimensions *Domain* and *Type* yield two combinations which occur frequently. First, *Address Data* with *Matching Data Vendor* (ranked 8th). These are most often vendors who offer to cleanse and improve the quality of customer databases. Second, *Social Media* with *Enrichment - Sentiment* (ranked 20th), which is explained by the fact that sentiment analysis is only meaningfully applied to text written by human beings, the main source of which is social media.

The 9th top result (*Data Origin: Community* and *Type: Data Market Place*) underlines the fact that data marketplaces are, in fact, similar to traditional marketplaces in the sense that ordinary people play an important role in them. This finding is also in line with the recent rise in popularity of crowdsourcing applications [4], which also rely on an active community.

Number 14 on the list is *Pre-Purchase Testability: Complete Access* and *Pricing Model: Free*. This co-occurrence is explained by the fact that free services always grant complete access before any purchases are made. Most of the surveyed vendors that fall into this combination are either authored by governments (similar to the data.gov movement), make their money with advertising or other third-party services, or are simply not in a competitive state (beta versions, research projects).

The final interesting finding is the high co-occurrence of *Data Output: Report* and *Type: Enrichment - Analysis* (ranked 17th). This suggests two things: First, analysis results are usually so complex that they need some kind of post-processing (e. g. visualization). Second, these results are most often targeted at decision makers, and not intended for further processing, which makes a report the ideal form for them.

Note that the other results, which have not been described in-depth, are most likely not interesting due to their obvious nature or because the high co-occurrence stems from the set-up of the dimensions (i. e. *Target Audience* and *Time Frame* are by design highly occurrent, because they only consist of two dimensions each).

3. Related Work

The work of GE et al. [5] can be seen as predecessors of data marketplace research. They investigated five different Question & Answer Web sites (e. g., Askjeeves.com) and studied their business models. However, no technical details have been analyzed. A more recent study has been performed by EDD DUMBILL [6], who takes a closer look at the four biggest data marketplaces (Infochimps, Factual, Windows Azure Data Marketplace, and DataMarket). However, the underlying methodology is not disclosed, the work is not reviewed and only published as a blog article.

A more organized approach was followed by MILLER who interviewed ten providers of data marketplaces or data related services in a series of podcasts [7]. However, he only provides the interviews in a rather unprocessed form, i. e., as audio files, which makes it difficult to access and aggregate the contained information. Later, he published a report [8] on data marketplaces and their business models, in which he identified common functionalities that data marketplaces offer, elaborated on potential business models and made some general predictions, such as increasing competition and a wider choice of data and sources.

Furthermore, there have been investigations into specific marketplaces, for instance on Kasabi [9], who went out of business in 2012; Freebase, who try to create a "collaboratively created graph database for structuring human knowledge" [10]; and Microsoft's Windows Azure Marketplace [11]. While this study is concerned with data markets that provide business data, also works exist that are concerned with data markets for personal data. For instance [12] describe how Facebook data can be of value to recommender systems or [13] who found that while people generally worry about their personal data, they are not willing to pay in order to protect or control their personal data.

4. Conclusion & Future Work

In a world where digitization and the trend to represent everything as digital data has become irreversible, it is not surprising that a new species of marketplaces, *data marketplaces*, is emerging. This study was the second iteration of an observational study of this new field of data vendors and data marketplaces. Together with its predecessor, this study is intended to establish a solid framework for categorization as well as provide initial data for an overview and early trend prediction. In the work reported here, we were able to come one step closer to this goal. Besides a description of the market situation in 2013, some forecasts regarding future developments could be made. However, it remains to be seen whether these predictions hold true; to this end, our study needs to be repeated once more in the not-too-distant future.

In summary, the following trends have been identified: Raw data became less frequently offered during the past year, while enrichment services and processed data are on the rise. This is evident by the numbers of vendors per type, but also by the fact that reports, specialized software, and Web interfaces were more often provided. Also supporting this assumption are the facts that data is available in many more languages and that the amount of up-to-date offerings increased. One reason behind this observation could be the fact that in this way, information can be used directly without much hassle. Regarding targeted customers, the numbers strongly suggest that data is a market which mainly focuses on business customers, while consumers are less relevant.

Although the domains in which data is offered have not changed tremendously, the origins of the data show an over-proportional growth in less established data sources such as self-generated, enhanced customer data, and community data. This may be owing to the fact that these sources are becoming more widely accepted (e. g., Wikipedia) or because these sources are the only way to obtain certain knowledge (for instance, some data is only of value if it can be well integrated with a company's existing data). Unfortunately, at this point we cannot say anything about the development of trustworthiness, since the data is too inconclusive.

The shift in pricing models away from freemium was somewhat surprising, while the rise in pay-per-use was rather expected. On a technical level, Web technologies overtook more traditional exchange formats. Whether these two observations are trends or just outliers remains to be seen in subsequent studies. Regarding the new dimensions *pre-purchase testability* and *pre-purchase information*, it can be stated that most vendors provide sufficient information for buyers to make educated decisions whether or not to buy a product or service.

The performed analysis of co-occurrences has confirmed the intuition that certain combinations across dimension categories make more sense than others, e. g., regarding data access and output, CSV files are offered as downloads, while XML-encoded data is usually accessible through an API. Looking at the market as a whole, it can be seen that the market is still in motion with four companies leaving the survey. However, at the same time we could observe a positive trend in company growth as well as a maturing tendency. Similar to the technical standards and the pricing models it remains to be seen how this develops further. Also, it should be kept in mind that one year is a rather short period in terms of business developments. From all that, it can be concluded that the market for data vendors is far from fully mature and leaves vast potential for development.

Acknowledgment. We thank Michael Glahn and Dennis Assenmacher for their support in conducting the survey.

References

[1] Muschalle A., Stahl F., Löser A., Vossen G. Pricing Approaches for Data Markets. In: *6th International Workshop on Business Intelligence for the Real Time Enterprise (BIRTE)*, 2012, 129–144.

[2] Schomm F., Stahl F., Vossen G. Marketplaces for data: an initial survey. *ACM SIGMOD Rec.*, 2013, 42(1), 15-26.

[3] Stahl F., Schomm F., Vossen G. The Data Marketplace Survey Revisited. *Technical Report No. 18, European Research Center for Information Systems*, 2014.

[4] Hammon L. Crowdsourcing: Eine Analyse der Antriebskräfte innerhalb der Crowd. *Schriftenreihe Innovative Betriebswirtschaftliche Forschung und Praxis*, Kovač, 2012.

[5] Ge W., Rothenberger M., Chen E. A Model for an Electronic Information Marketplace. *Australasian Journal of Information Systems*, 2005, 13(1).

[6] Dumbill E. Data markets compared [WWW] http://strata.oreilly.com/2012/03/data-markets-survey.html (accessed 10.04.2014)

[7] Miller P. Data Market Chat [WWW] http://cloudofdata.com/category/podcast/data-market-chat/ (accessed 10.04.2014)

[8] Miller P. Data markets: in search of new business models [WWW] http://pro.gigaom.com/2012/08/data-markets-in-search-of-new-business-models/ (accessed 10.04.2014)

[9] Möller K., Dodds L. The Kasabi Information Marketplace. In: *21st World Wide Web Conference, Lyon, France*, 2012.

[10] Bollacker K., Evans C., Paritosh P., Sturge T., Taylor J. Freebase: a collaboratively created graph database for structuring human knowledge. In: *Proc. ACM SIGMOD*, New York, USA, 2008, 1247–1250

[11] Microsoft Whitepaper: Windows Azure Marketplace [WWW] http://go.microsoft.com/fwlink/?LinkID=201129&clcid=0x409 (accessed 10.04.2014)

[12] Gottschlich J., Heimbach I., Hinz O. The Value Of Users' Facebook Profile Data - Generating Product Recommendations For Online Social Shopping Sites. In: *ECIS*, 2013.

[13] Potoglou D., Patil S., Gijón C., Palacios J. F., Feijóo C. The value of personal information online: Results from three stated preference discrete choice experiments in the UK. In: *ECIS*, 2013.