Marketplaces for Data: An Initial Survey

Luizooooooo

LM Informatica

A.A. 2024/2025

Introduction

- Despite the sheer quantities of data available on the web, such information is not easy to find

- In response to the demand, a new category of professionals emerged who took on the function of information intermediaries (for a fee)

- Commonly they process, sell and resell data available on the web

- Numerous new forms of marketplaces for data have emerged

Introduction: Survey details

- We have conducted a comprehensive survey and comparison of multiple data marketplaces and data vendors

- The survey includes 46 suppliers of data, conducted from April to July 2012

- Aim of identifying categories and dimensions of data marketplaces as well as vendors of data in order to build a taxonomy for data marketplaces

- Researching the market and its developments, in particular

  - Understand market dynamics

  - New research opportunities regarding the application of new technologies

Methodology

The survey focuses on online web services that offer either:
- Trading data
- Raw data
- Data enrichment tools

We distinguish the following services:
- **Data marketplace**
  Platform on which anybody can upload and maintain data sets
- **Data vendor**
  Someone who has data and offers it to others, either for a given fee or free of charge
- **Data enrichment services**
  Service that take input from the user and enhance it in some way, eg. by analyzing or tagging data

Approach

We classified the marketplaces based on *The Kasabi Information Marketplace*[1] and expanded to **12 boolean dimensions** of categorized vendors.

Table 1: Set of dimensions.

	Dimension	Categories	Question to be answered
objective	Type	Web Crawler, Customizable Crawler, Search Engine, Pure Data Vendor, Complex Data Vendor, Matching Vendor, Enrichment Tagging, Enrichment Sentiment, Enrichment Analysis, Data Market Place	What is the type of the core offering?
	Time Frame	Static/Factual, Up To Date	Is the data static or real-time?
	Domain	All, Finance/Economy, Bio Medicine, Social Media, Geo Data, Address Data	What is the data about?
	Data Origin	Internet, Self-Generated, User, Community, Government, Authority	Where does the data come from? Who is the author?
	Pricing Model	Free, Freemium, Pay-Per-Use, Flat Rate	Is the offer free, pay-per-use or usable with a flat rate?
	Data Access	API, Download, Specialized Software, Web Interface	What technical means are offered to access the data?
	Data Output	XML, CSV/XLS, JSON, RDF, Report	In what way is the data formatted for the user?
	Language	English, German, More	What is the language of the website? Does it differ from the language of the data?
	Target Audience	Business, Customer	Towards whom is the product geared?
subjective	Trustworthiness	Low, Medium, High	How trustworthy is the vendor? Can the original data source be tracked or verified?
	Size of Vendor	Startup, Medium, Big, Global Player	How big is the vendor?
	Maturity	Research Project, Beta, Medium, High	Is the product still in beta or already established?

Limitations

- The facts about the data vendors were gathered by means of a Web search.

- The information we used was taken directly from the website of each vendor.

- The market of data vendors and data marketplaces is highly active. Therefore, it cannot be guaranteed that this study is fully exhaustive with regard to the number of vendors in the market.

- We are aware of the fact that a certain amount of data is traded directly between (large) corporations or within a certain ecosystem (such as social networks) without the use of intermediaries.

2013 Revisitation

- In 2014, another paper was published by the name of *The Data Marketplace Survey Revisited*[2]

- The same researches take into study the changes in the marketplace with the same research approach over the course of summer 2013

- After each dimension, we will see the changes reported in the follow up paper

- We'll also take a look at the potential reasons for the reported changes

Findings: Type

What the core of the product is
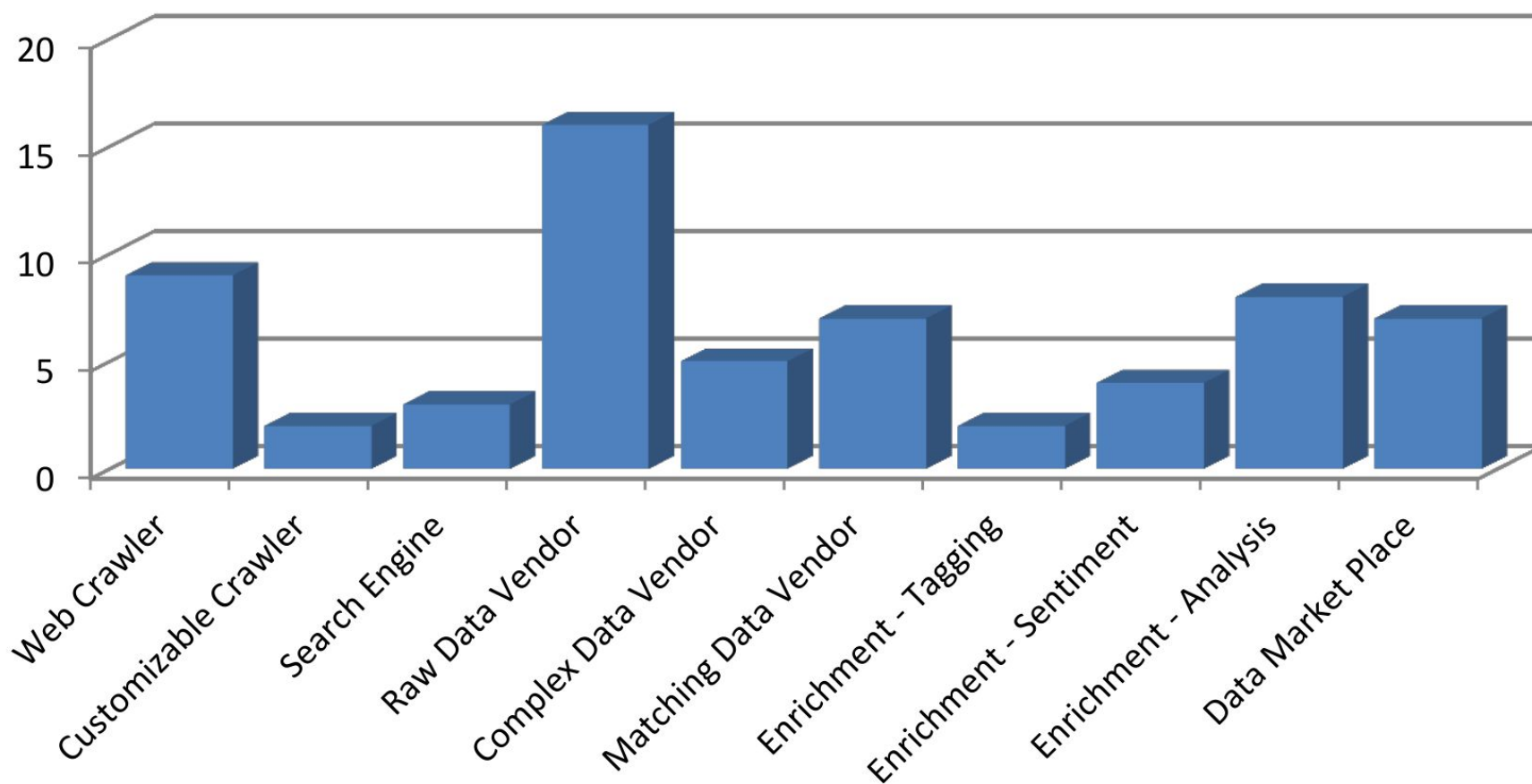
(These categories are not mutually exclusive and one vendor can fulfill the criteria of multiple categories)

Revisitation: Type

- Products that offer enhanced data are increasingly common
- The number of services for unprocessed information decreased slightly.
- Increasing demand in processed data
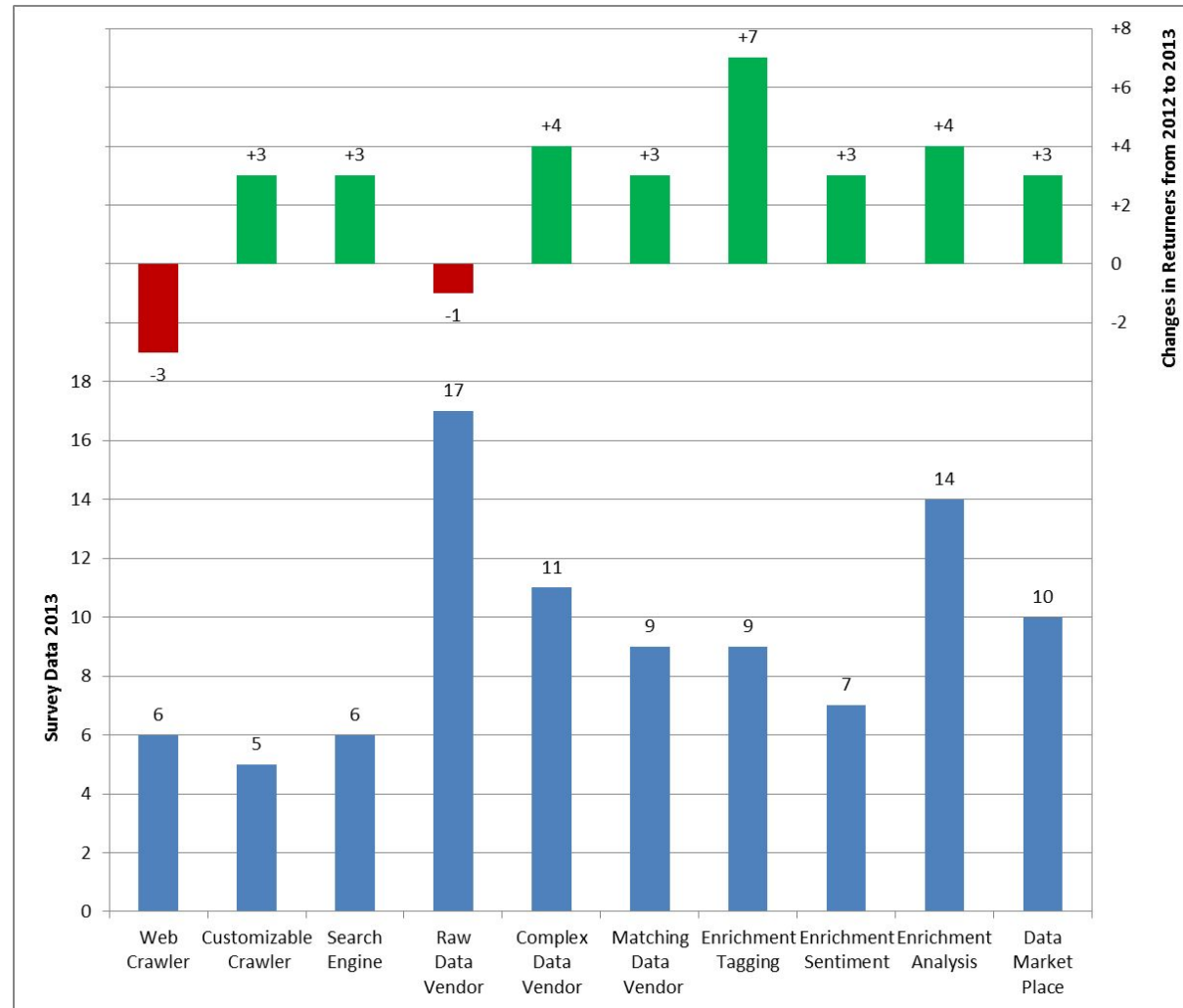
Findings: Time Frame

Captures the temporal context of the data.

- **Static/Factual:** Data is valid and relevant for a long period of time and does not change abruptly (eg. population numbers)
- **Up To Date:** Data is important shortly after its creation and loses relevance quickly (eg. social media entries)
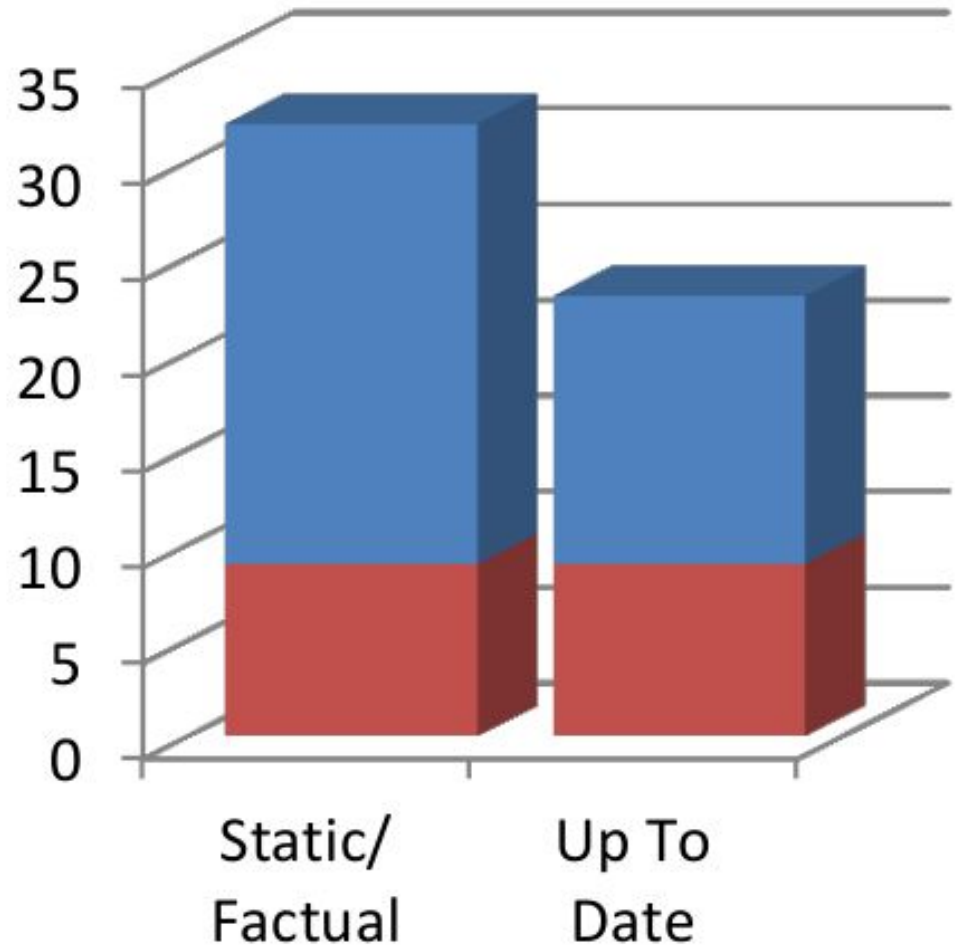
Revisitation: Time Frame

- **Stronger increase in up-to-date information within the returners group**
- **May caused from more "up to date" services in the industry (eg. socials)**
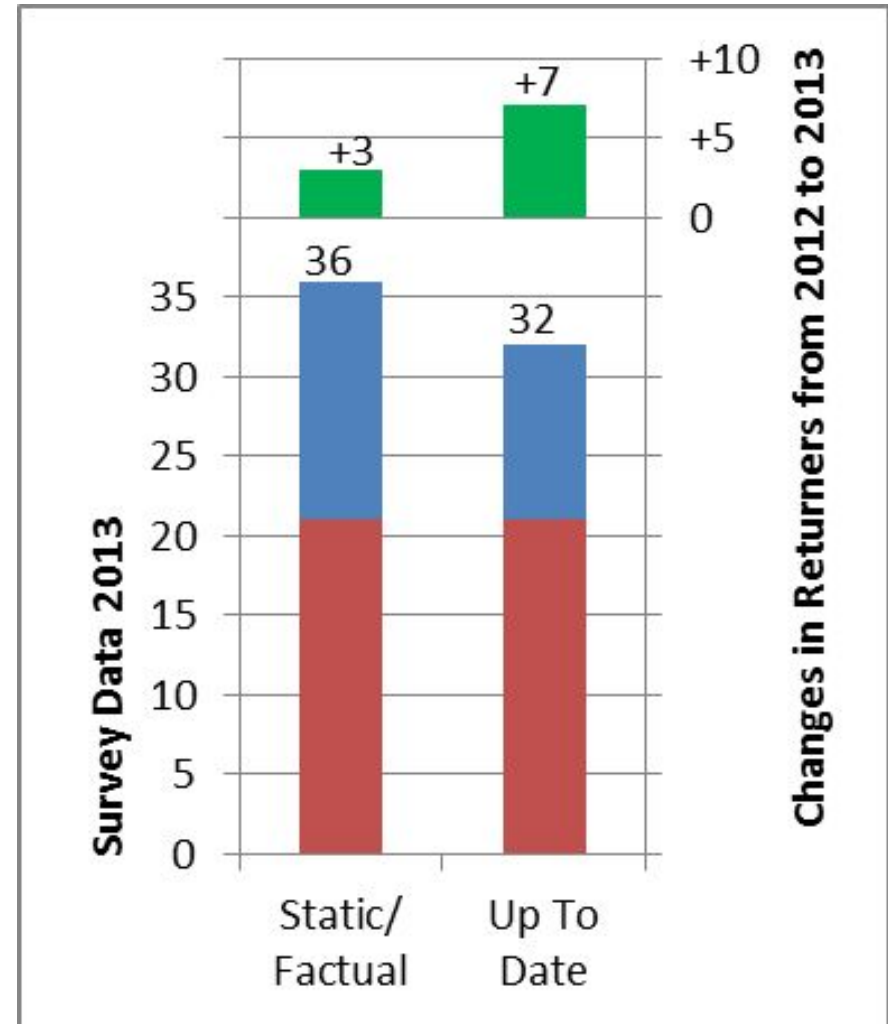
Findings: Domain

What the actual data is about
- **Any** means vendors that are not restricted and can incorporate arbitrary domain
- They include marketplaces, search engines, crawlers,... They serve the domain based on needs and uploads.
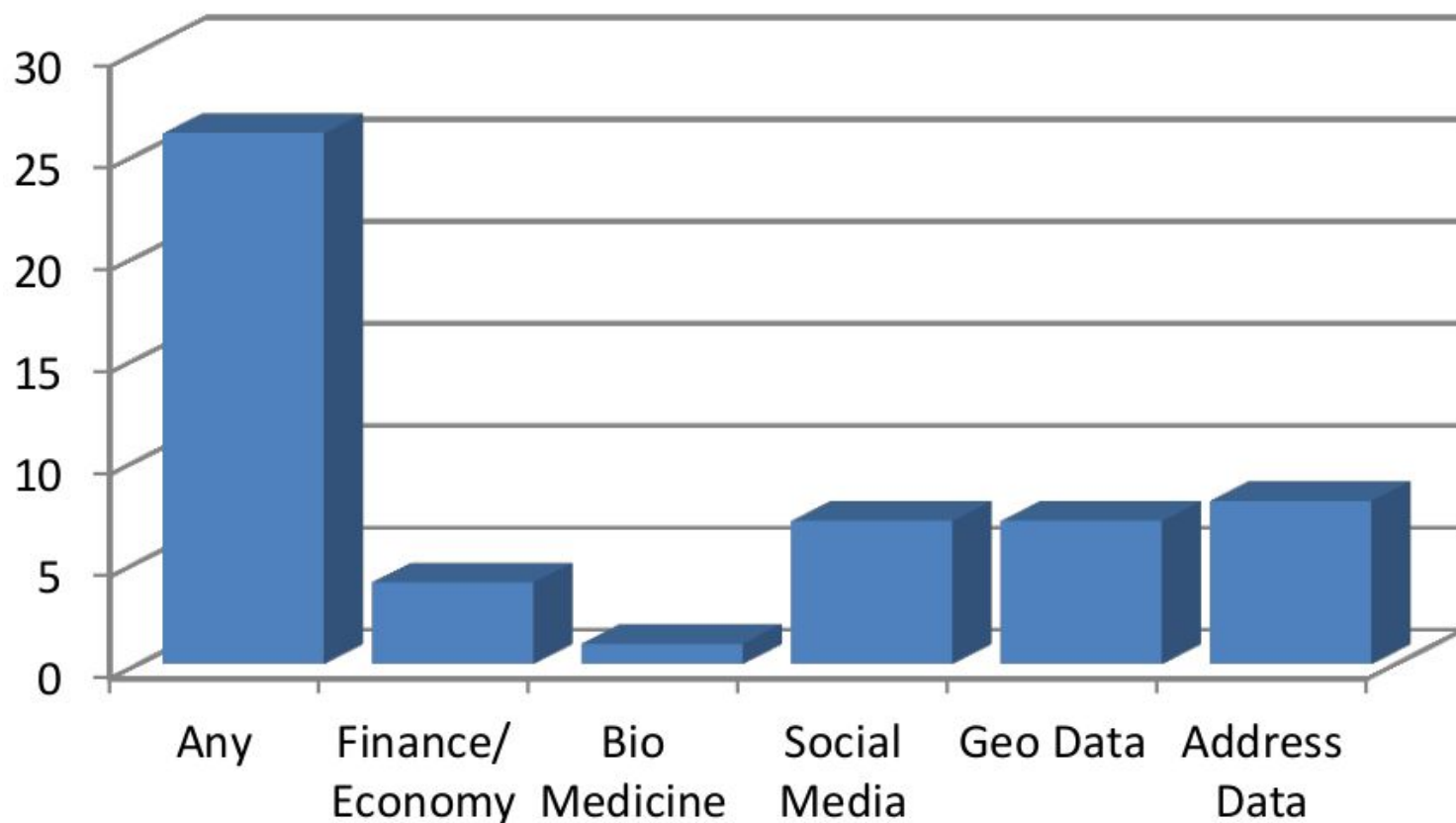
Revisitation: Domain
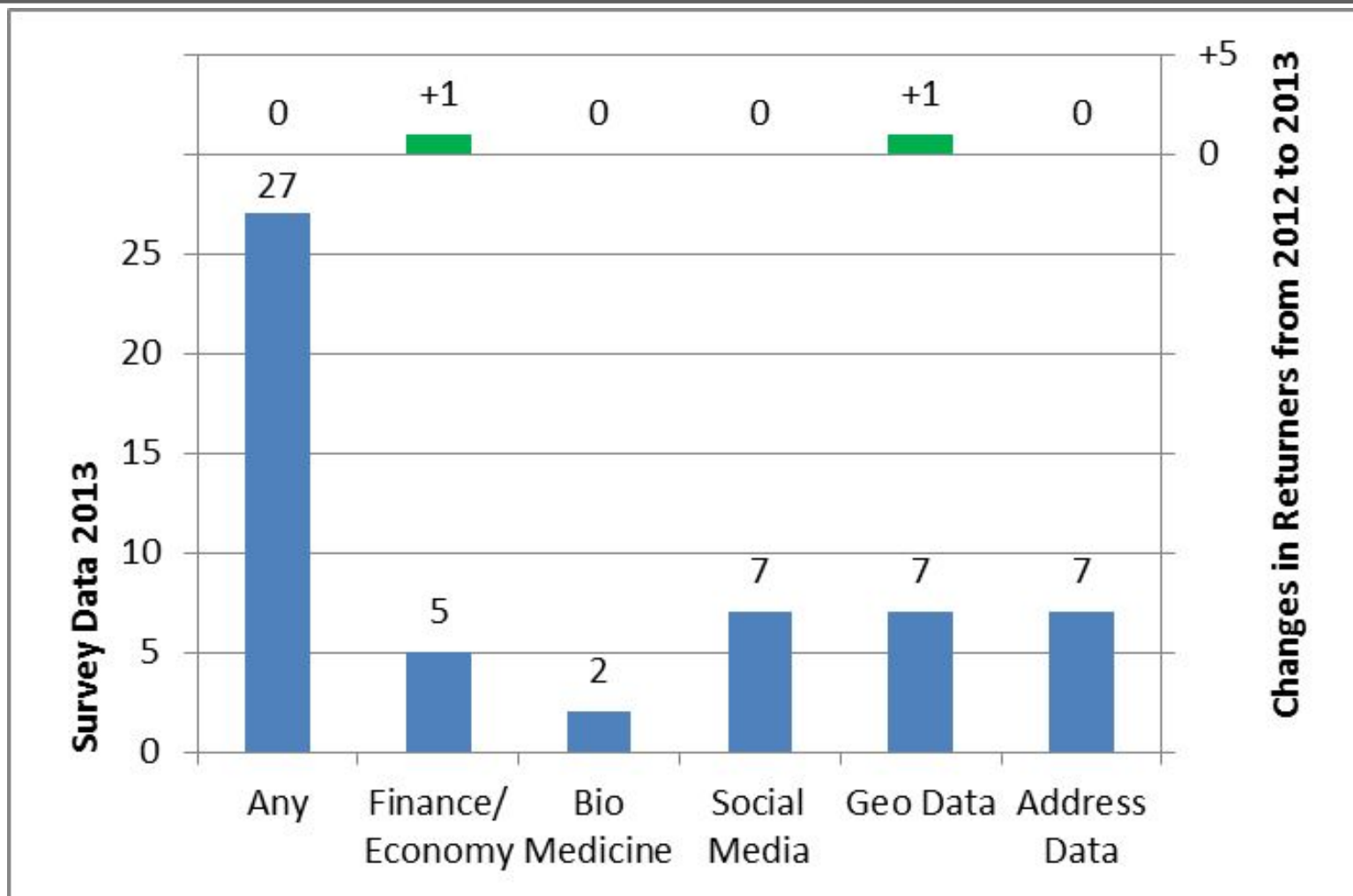
No change emerged in the revisitation under this dimension.

Findings: Data origin
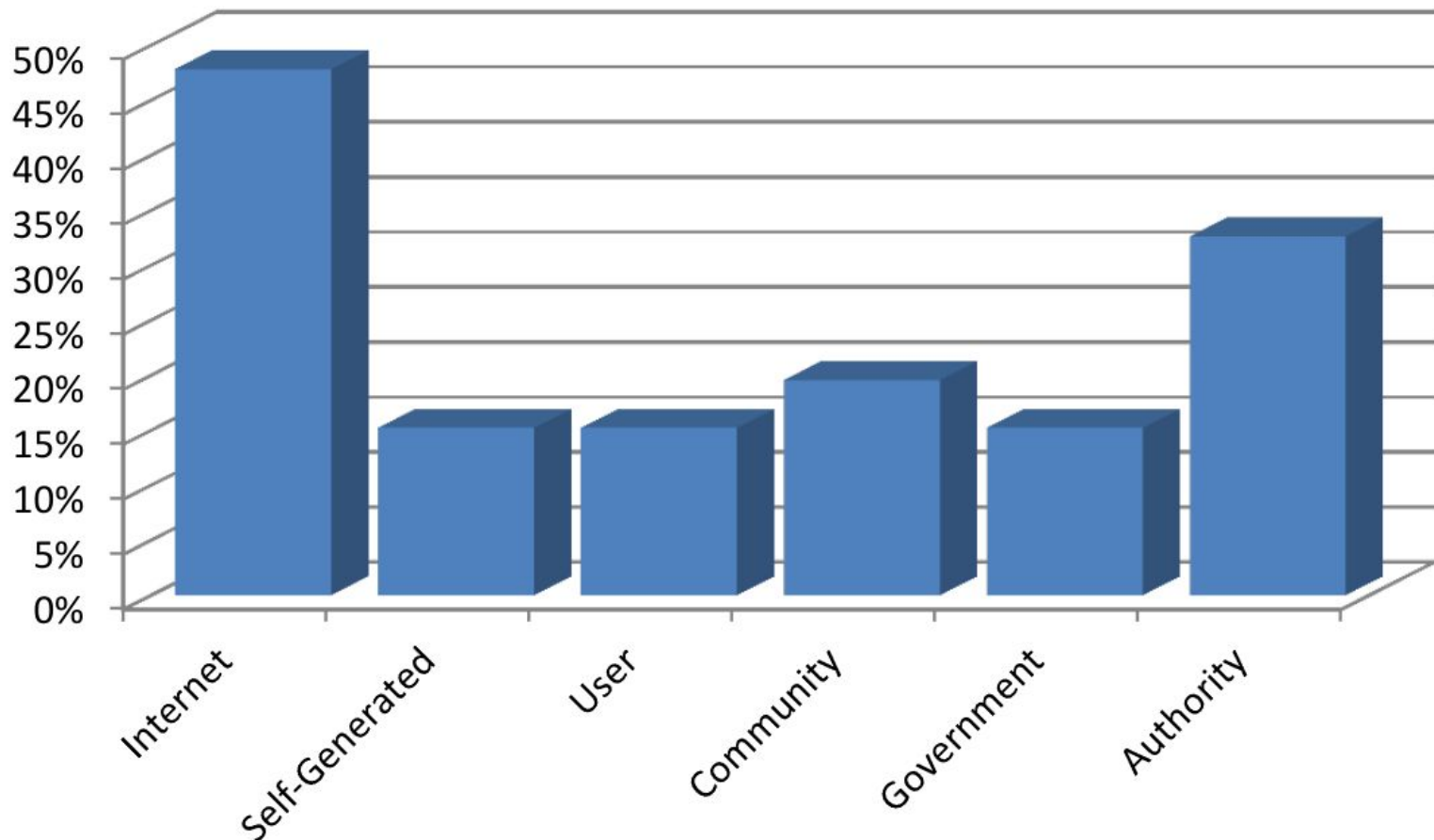
Where data comes from
- The majority of data originates from the internet and authoritative sources

Revisitation: Data origin

- 80% increase was in Self-generated, user and community origins.
- May indicate an increasing need in data that cannot be obtained through trusted sources.
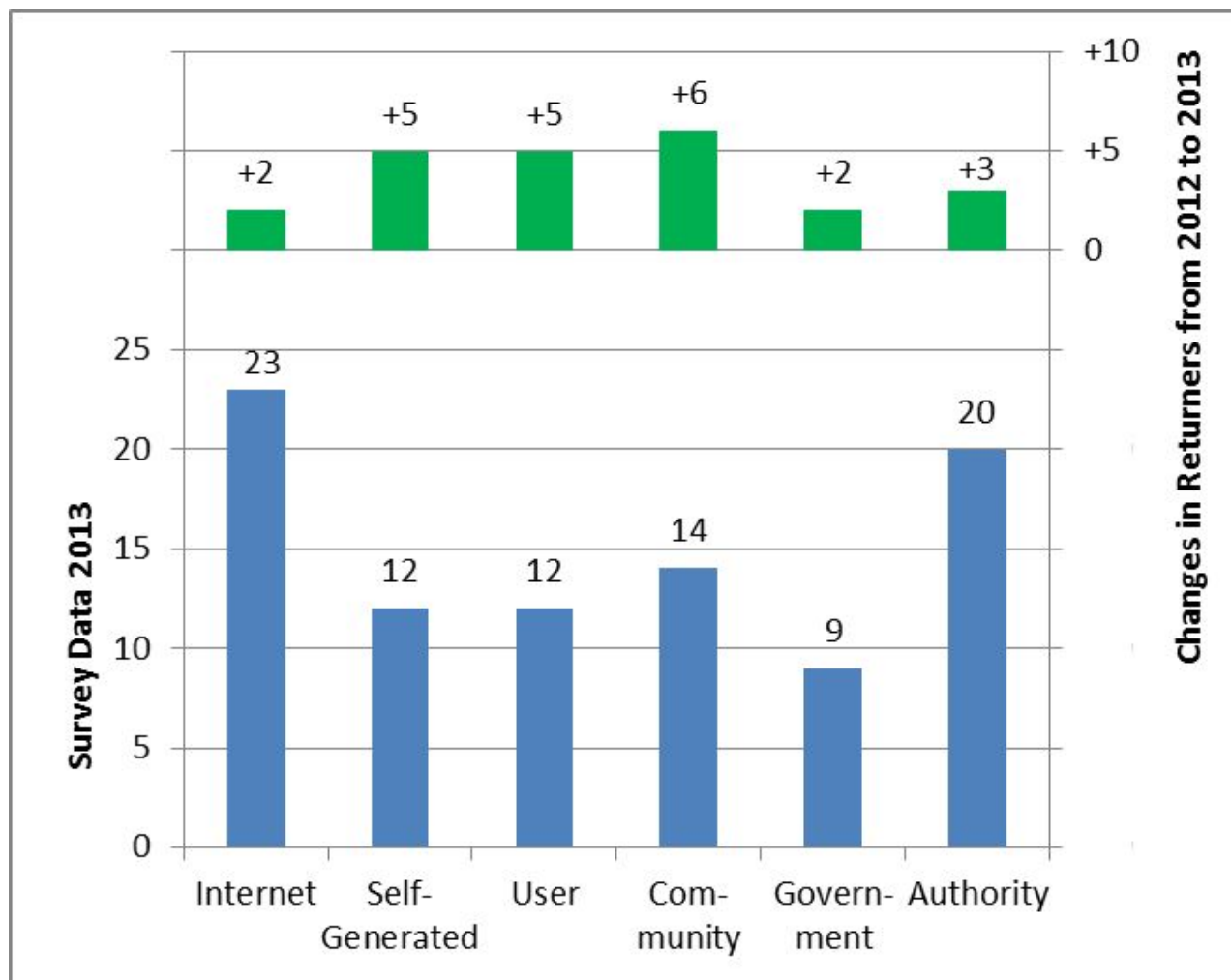
Findings: Pricing model
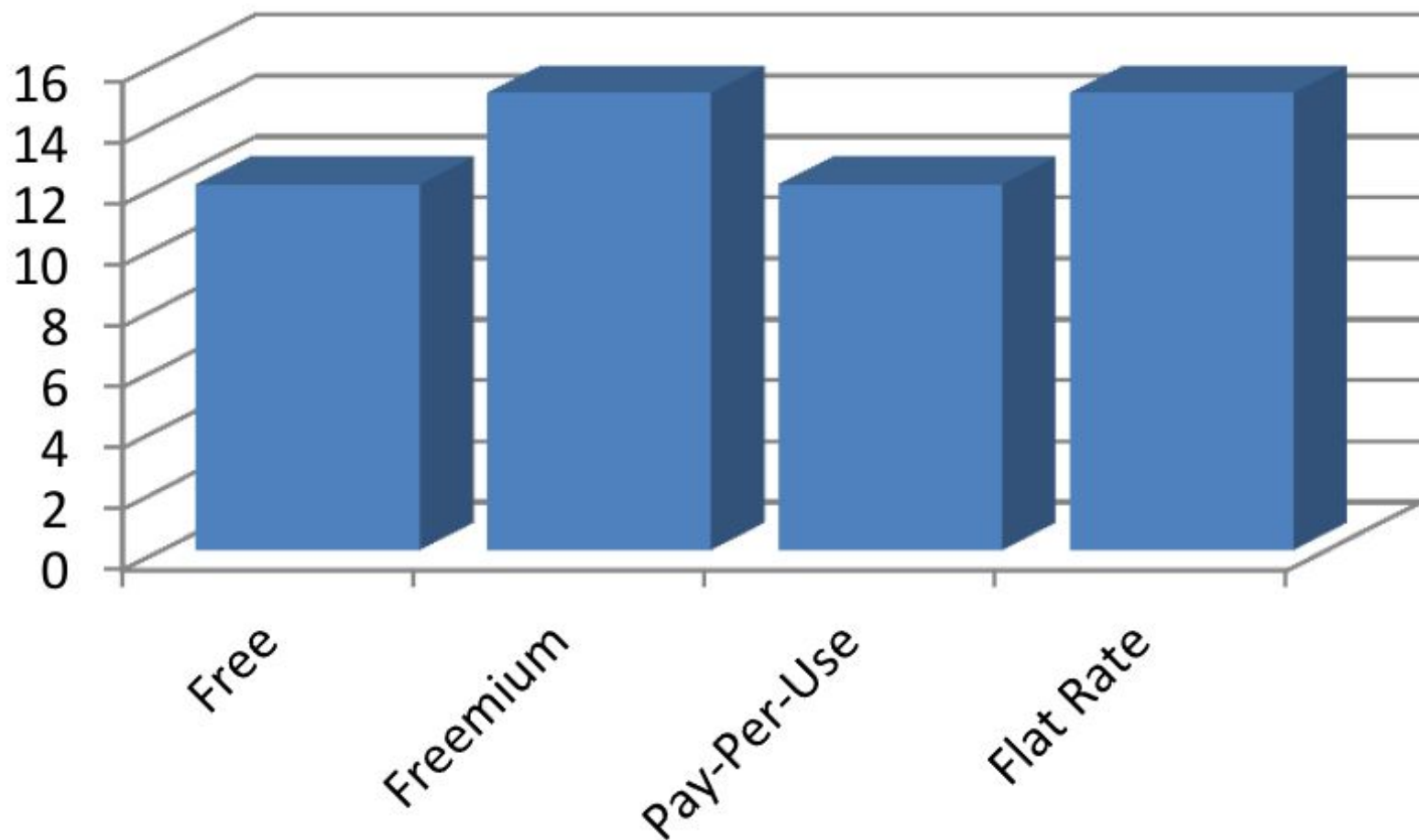
Business model divided in 4 main models
● Many vendors overlap due to combined pricing models

Revisitation: Pricing model

- Freemium lost importance
- Pay-per-use increased
- Customers may have more trust in the quality of the purchased data sets, thus higher willingness to pay.
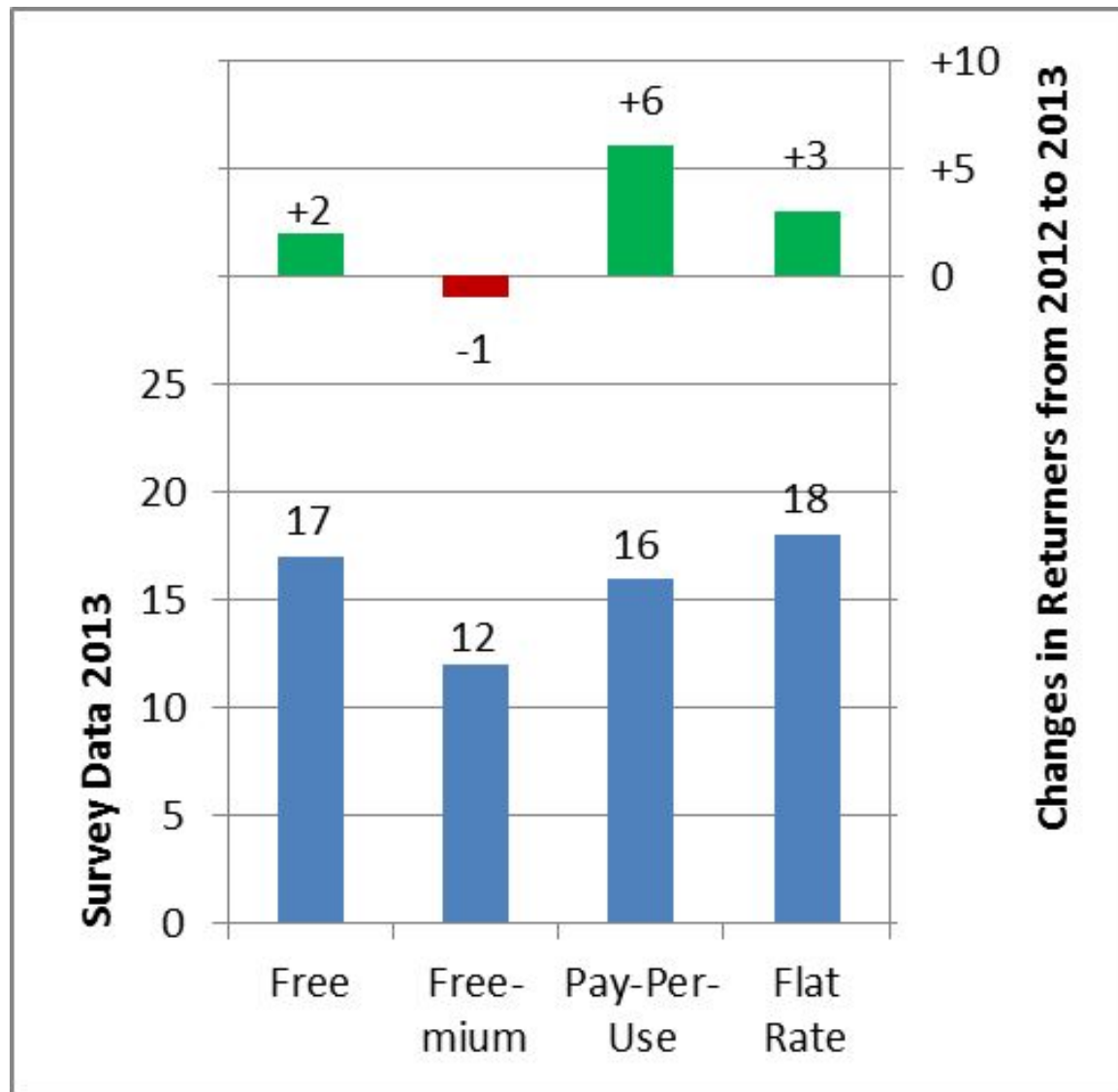
Findings: Data access

How end-users receive their data from vendors
- API is flexible, less than 30% of vendors have only API as a means to retrieve data
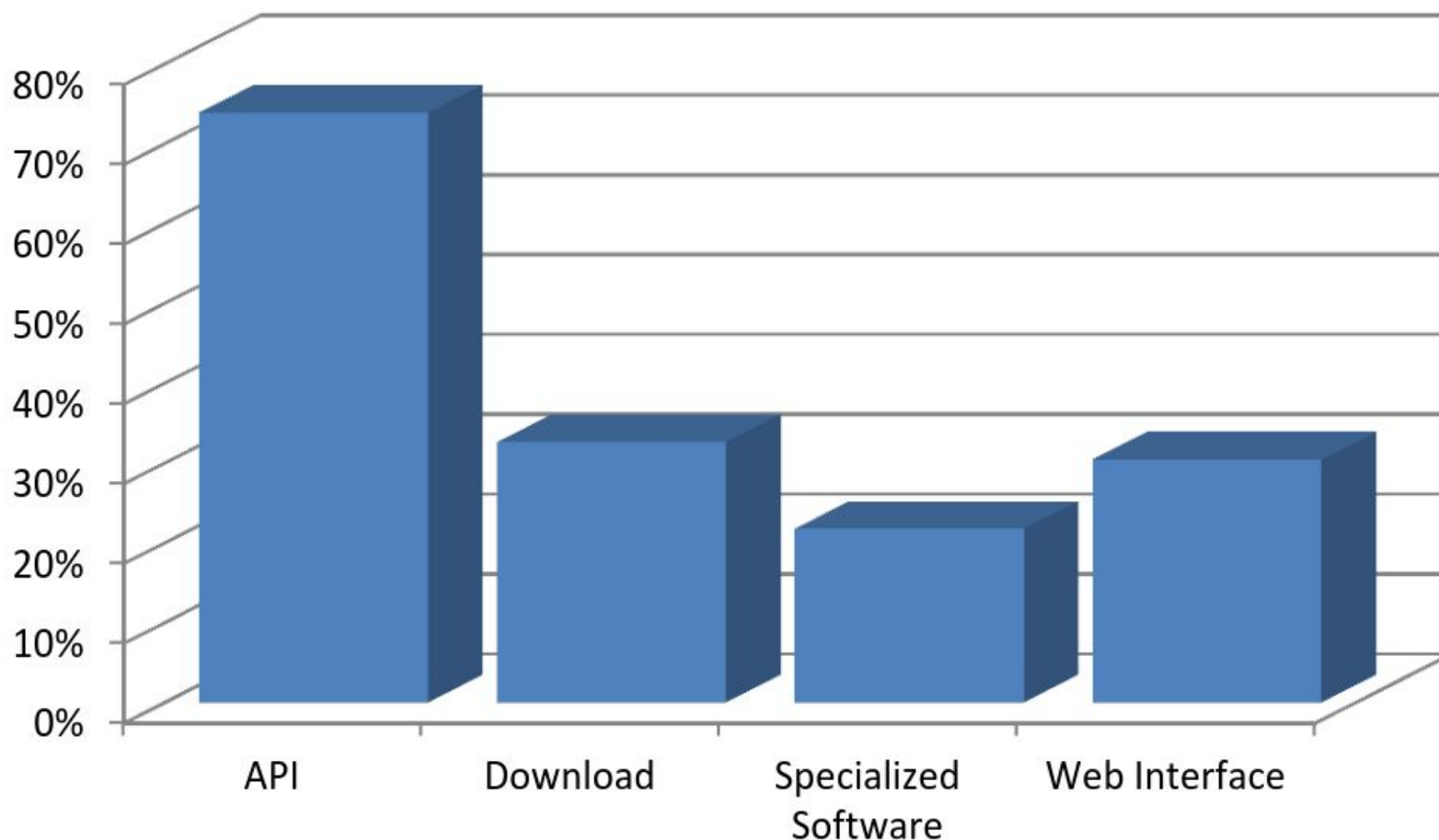
Revisitation: Data access

- specialized software has the strongest increase with about 60% for the returners group
- This could potentially stem from efforts to differentiate a company's offer from that of the competition
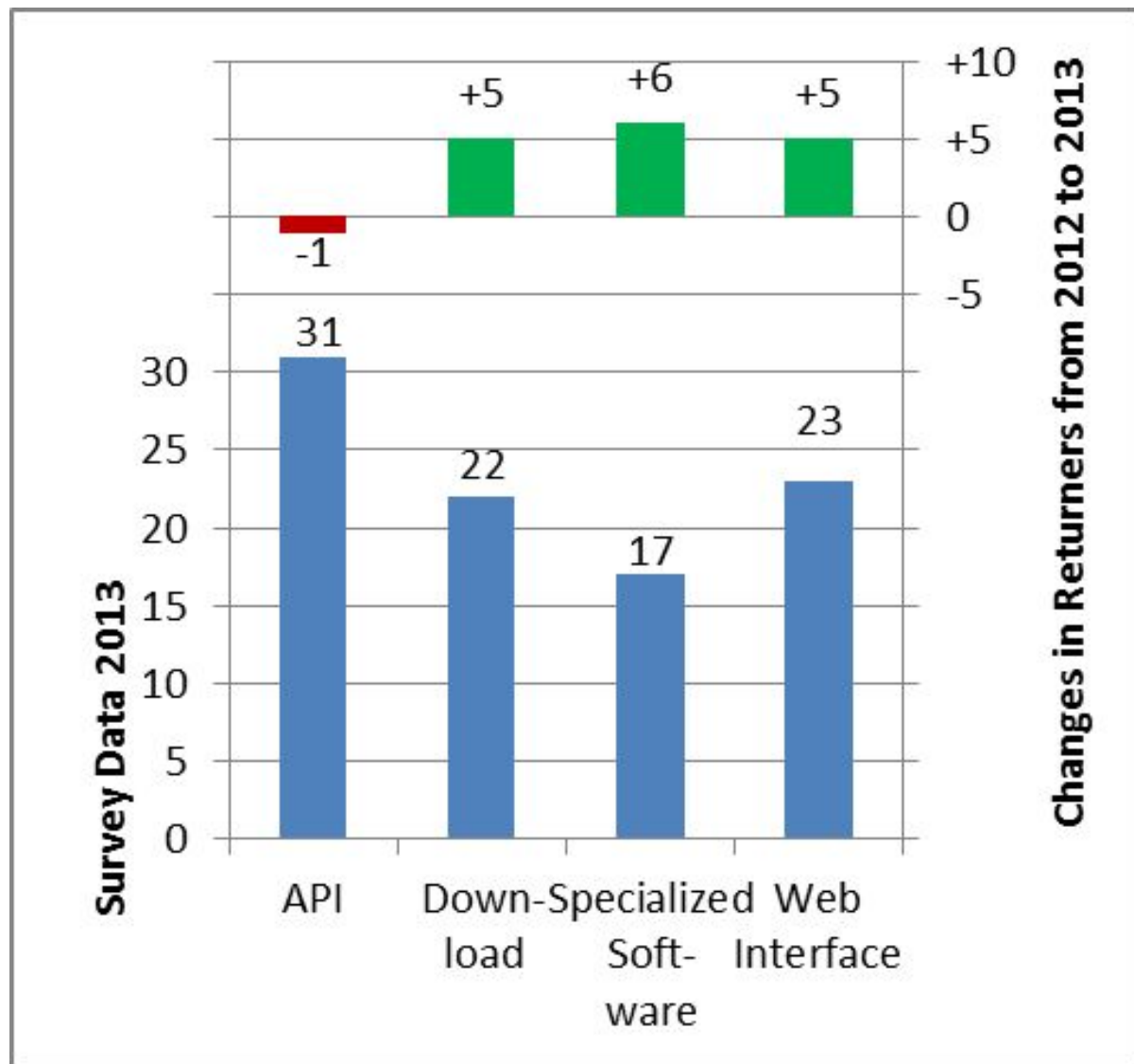
Findings: Data output

Given format of data
- CSV/XSL is the most popular category, but rarely it is the only one.
- Consistency with API data access type (many choose JSON/XML)
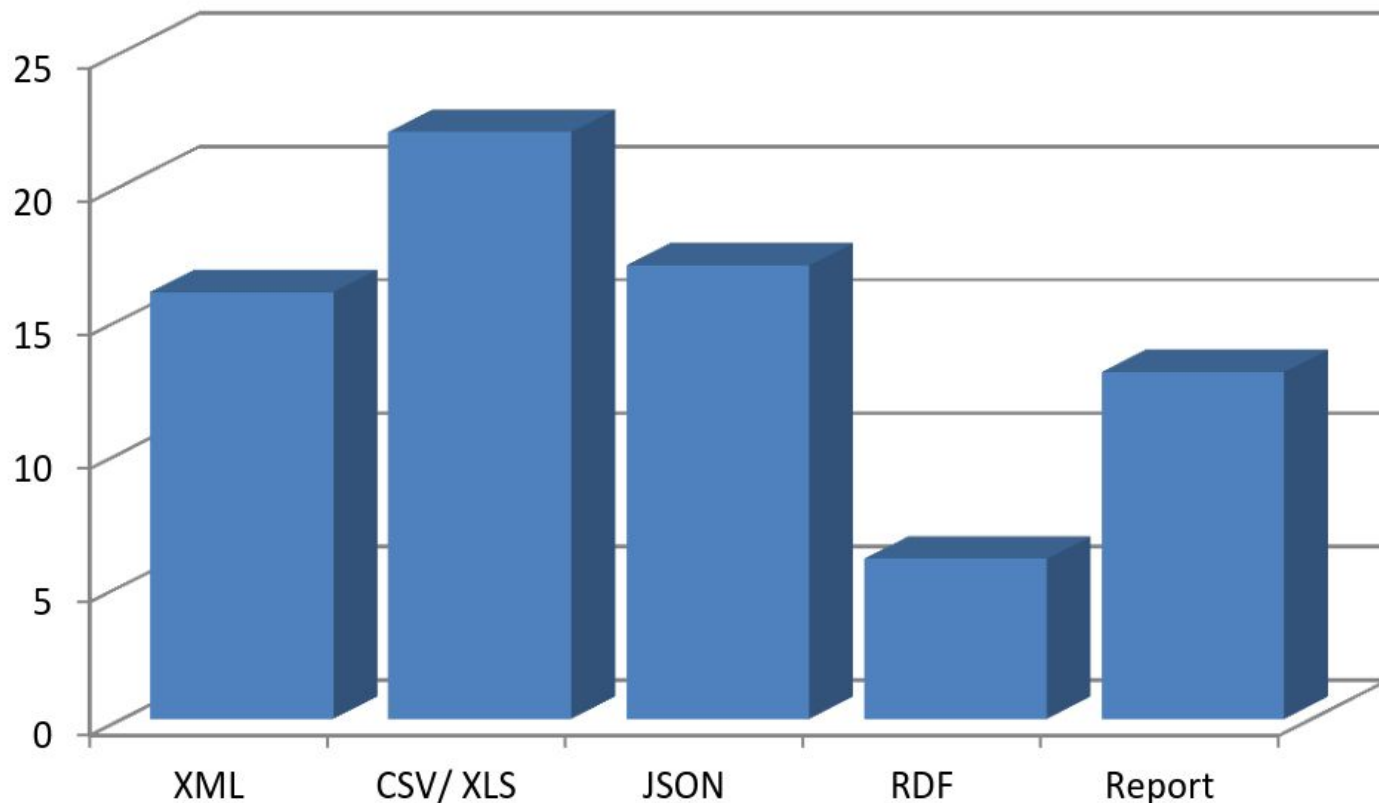
Revisitation: Data output

- Web standards could be about to replace the traditional exchange formats
- The increase in pre-formatted reports is feeding the impression from previous sections that vendors try to individualize themselves
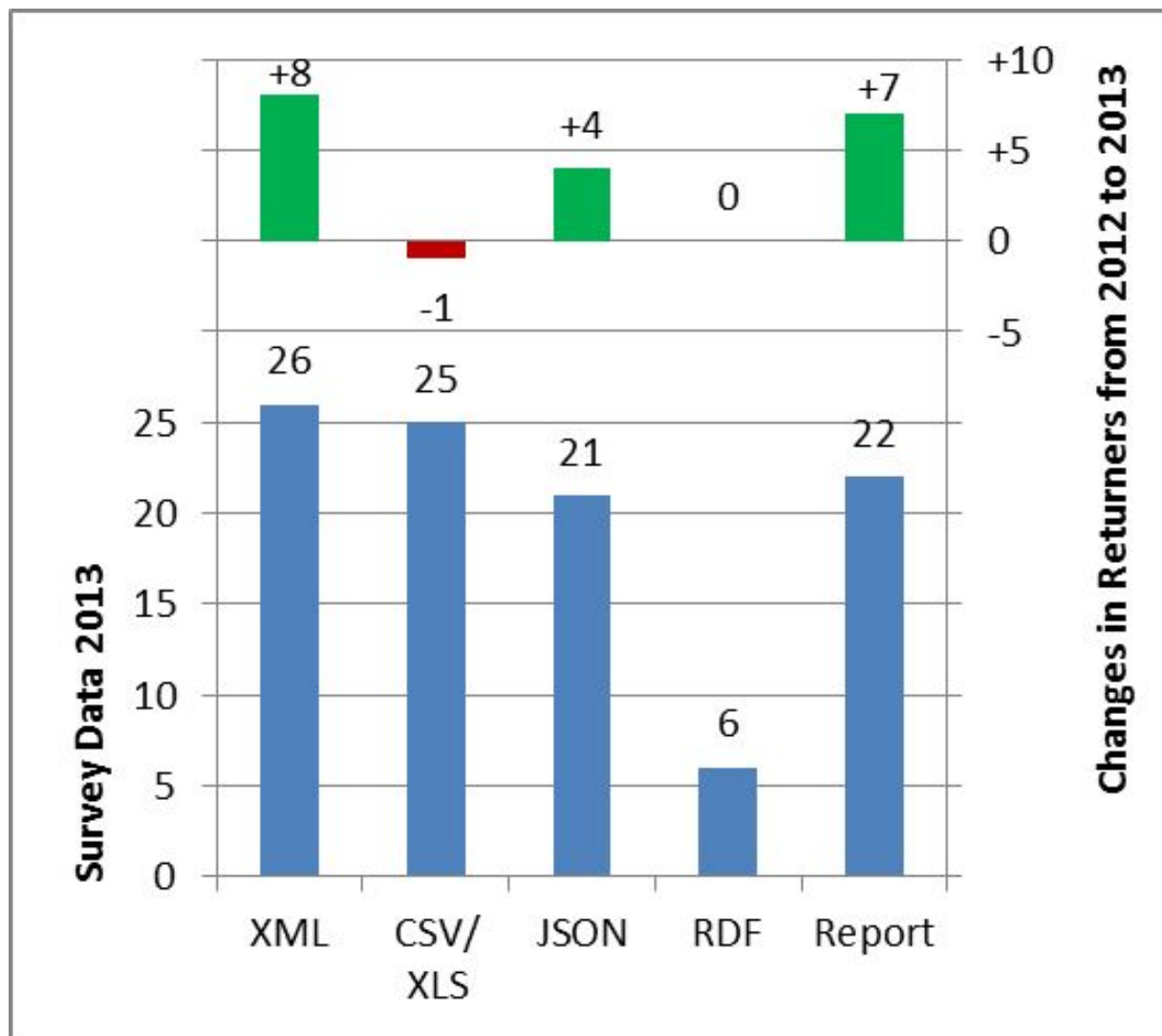
Findings: Language

Language of the website (left) and data (right)
- 98% of vendors use english
- Researches accounted only for English and Deutsch because of personal language skills.
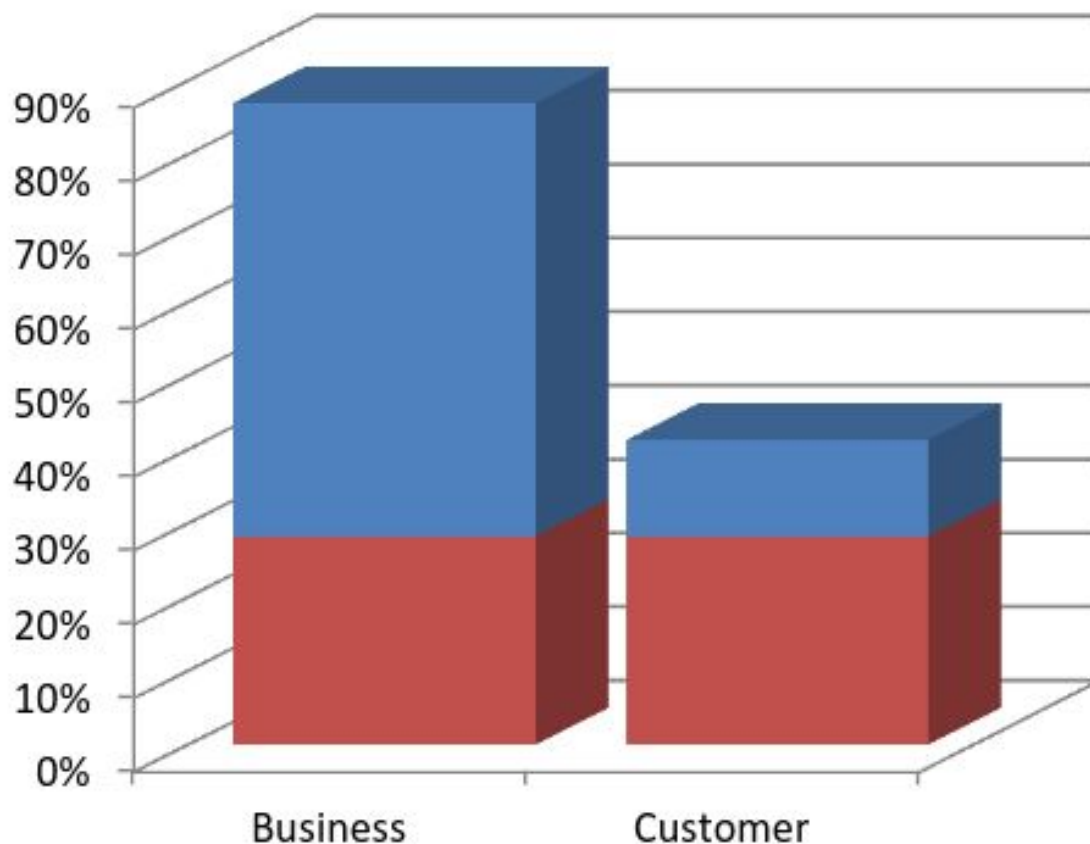- More languages are supported by bigger companies (eg. Microsoft)
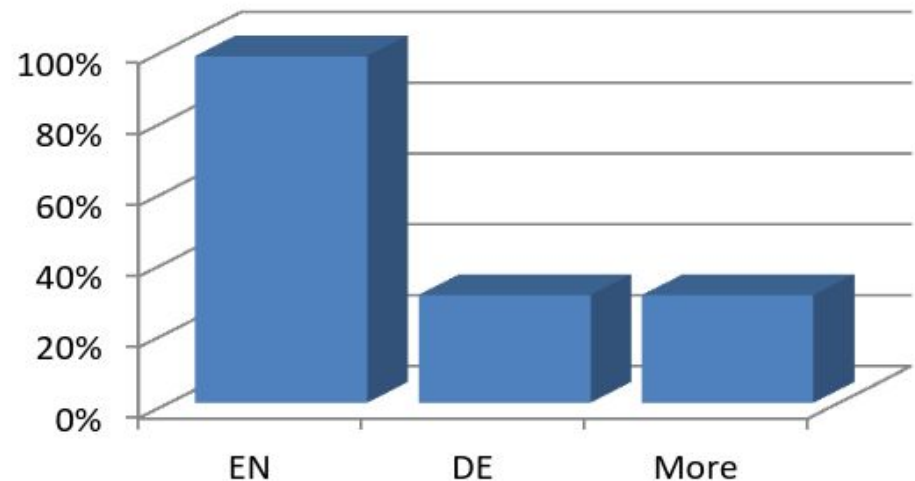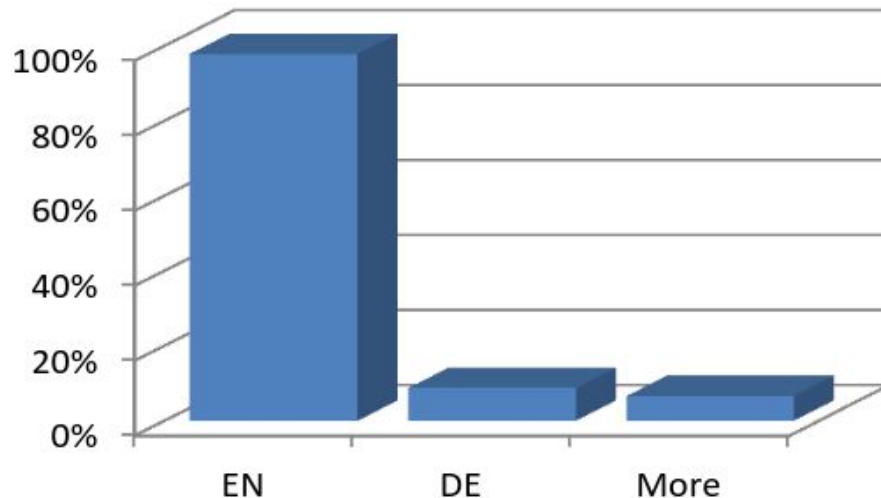
Revisitation: Language

- Non-english data became much more common
- This suggests that there is a high demand for national, non-English data.
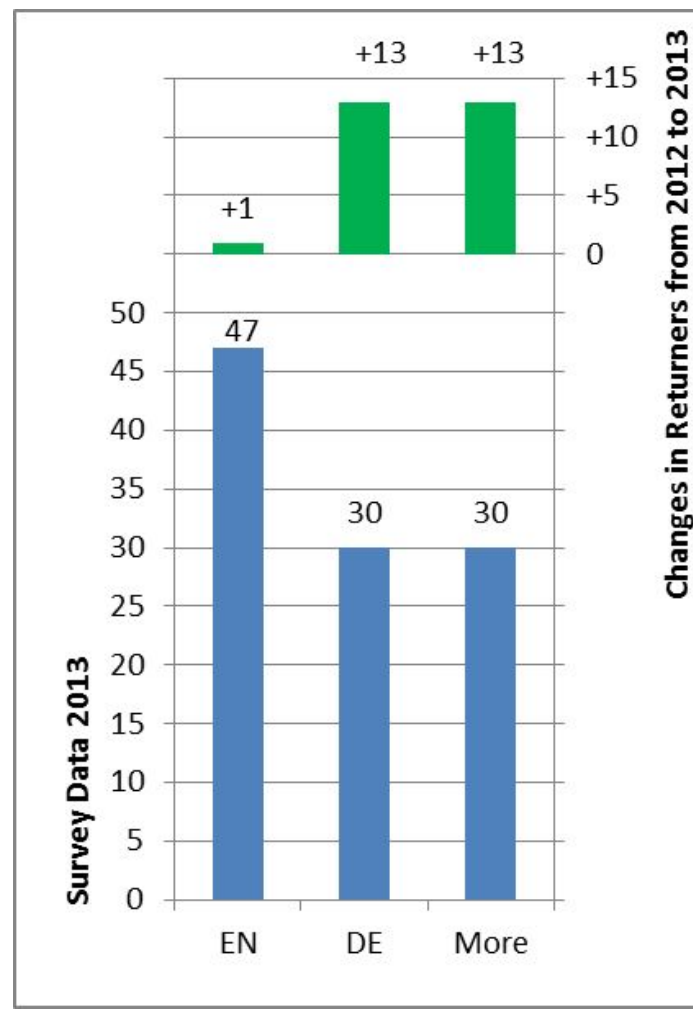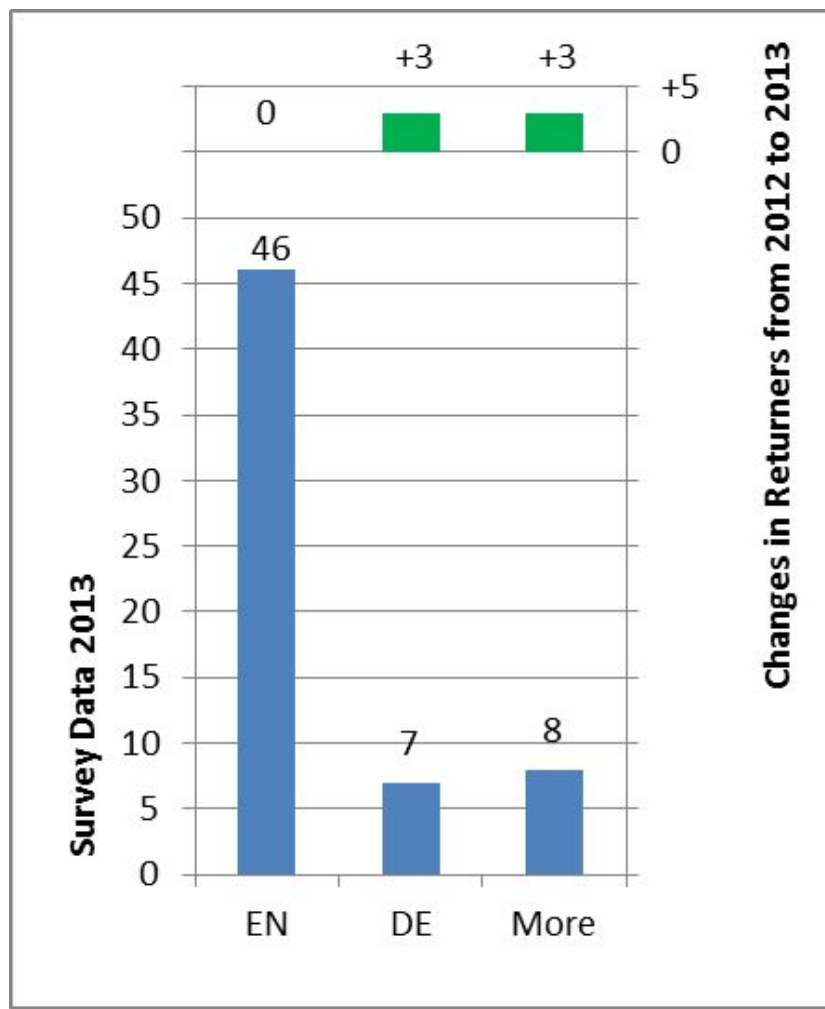
Findings: Target audience

Whom offerings are tailored towards.
- More general vendors target both audiences (eg. Kasabi, Windows Azure marketplace)
- Many choose to focus on businesses (87%)

Revisitation: Target audience

- Trend continues to favor business customers

- Considering the returners deltas between the study of 2012 and 2013, it seems reasonable to conclude that data services currently are and most likely will remain a B2B-centric market.
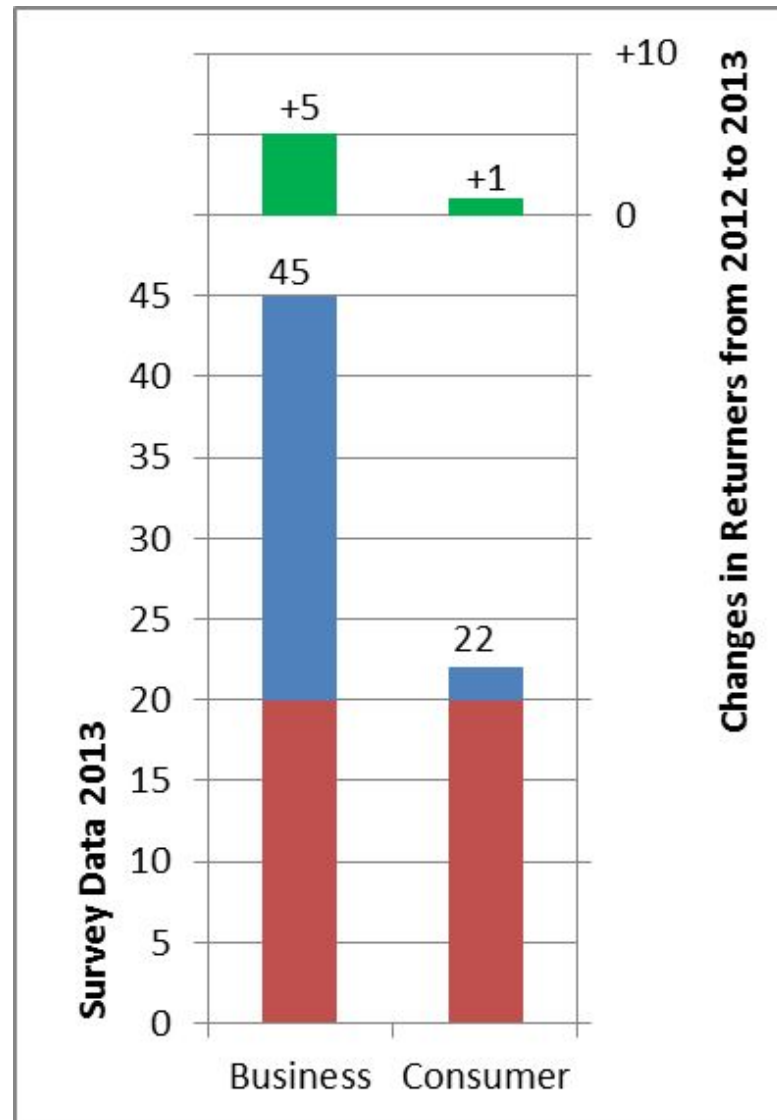
Findings: Trustworthiness

Subjective dimension, depending on the origin of the data as well as on how it is processed.

- For instance, data that come from a community could have a lower trustworthiness than data that is sourced from an authority.
- Typical vendors in the Low category are those that do not even claim to deliver correct or complete data
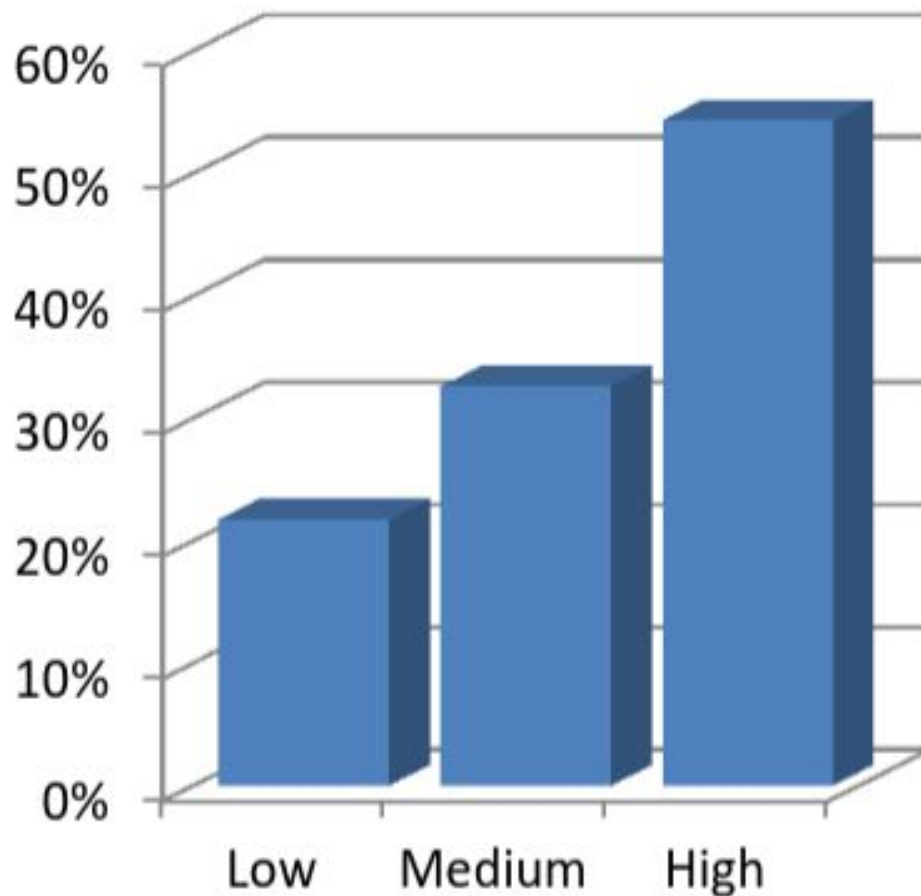
Revisitation: Trustworthiness

- There is no clear trend recognizable.

- While an increase on both ends can be observed, at the moment any interpretation of that phenomenon would be sheer speculation.
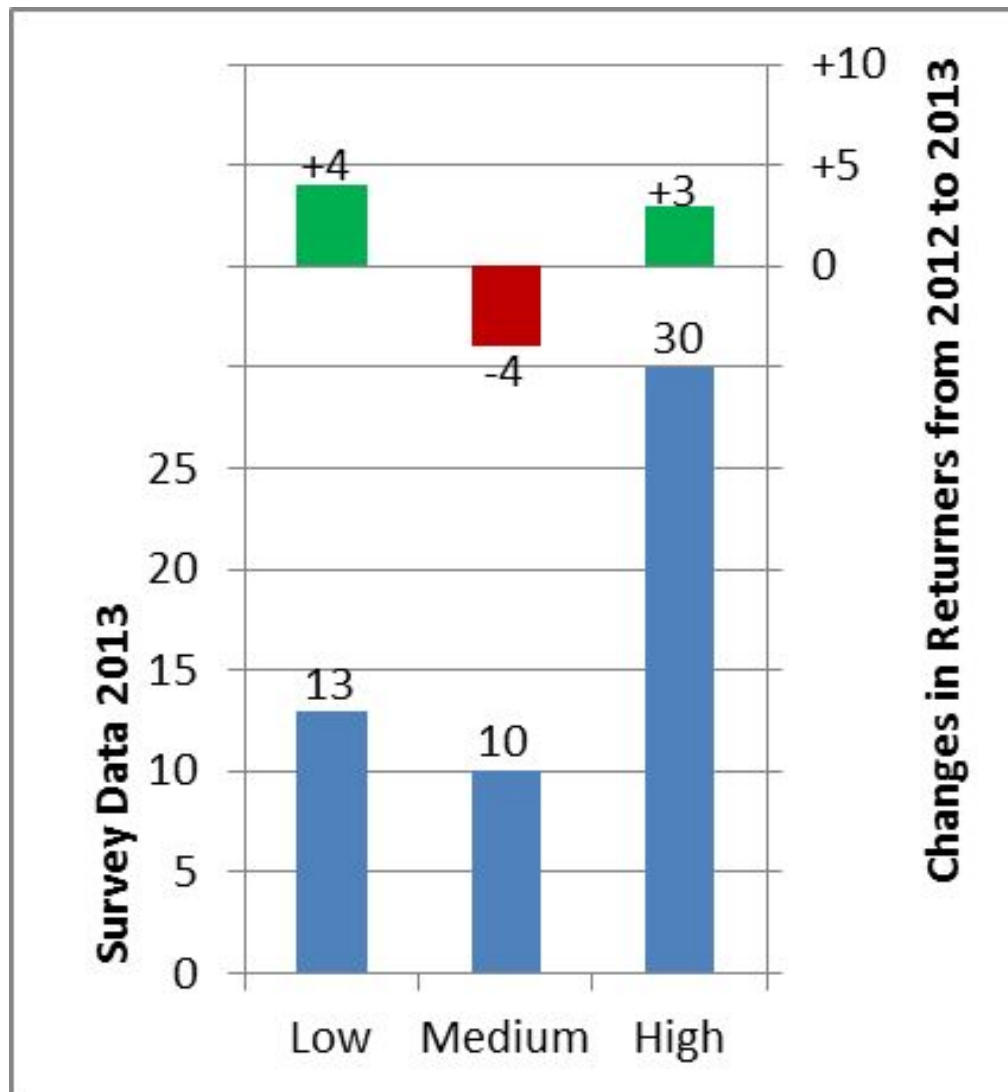
Findings: Size of vendors

Subjective dimension of vendors divided from startup to global player

- Global player examples: Yahoo!, Microsoft, IBM, etc.
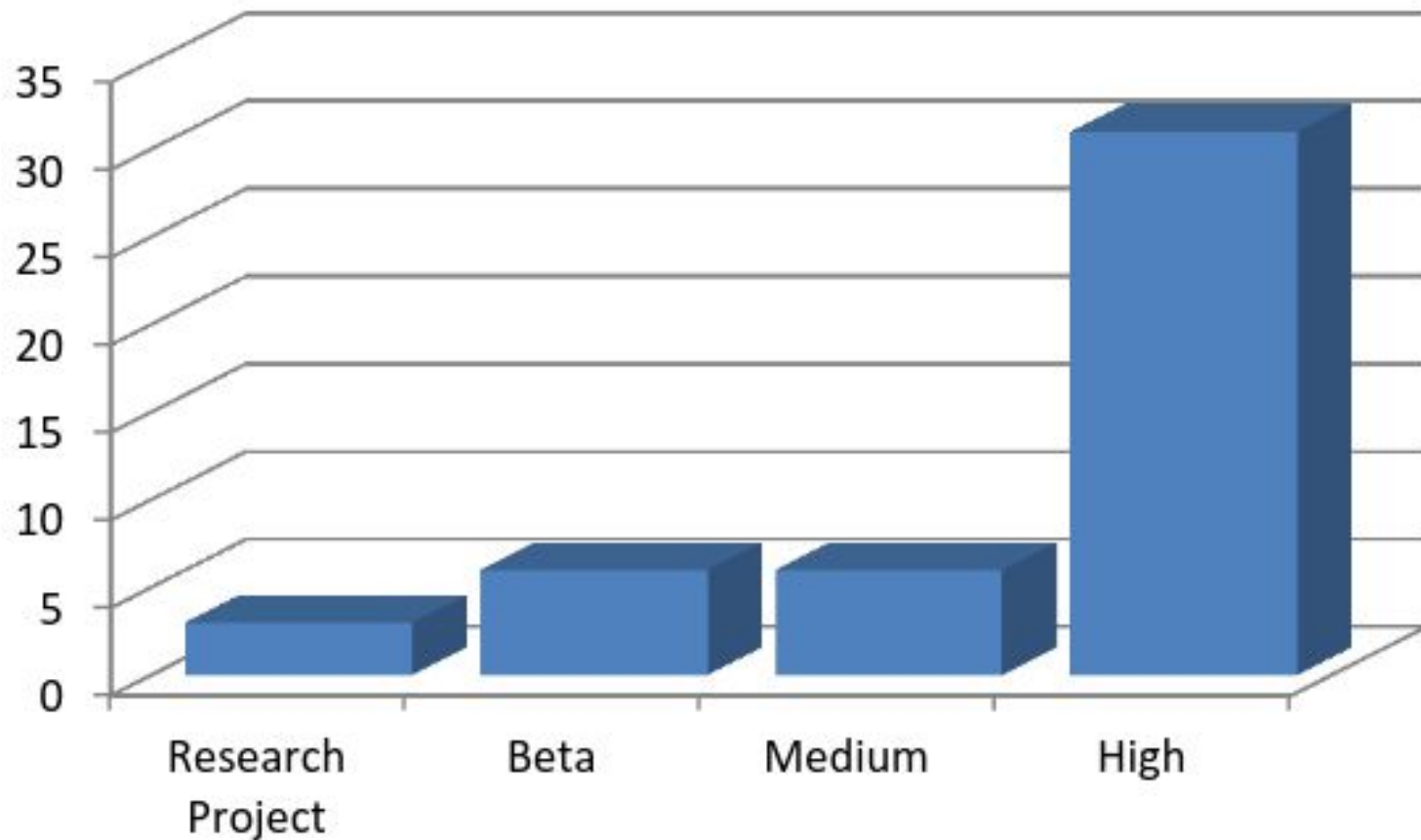- A low startup count indicates that the market is not easy to enter
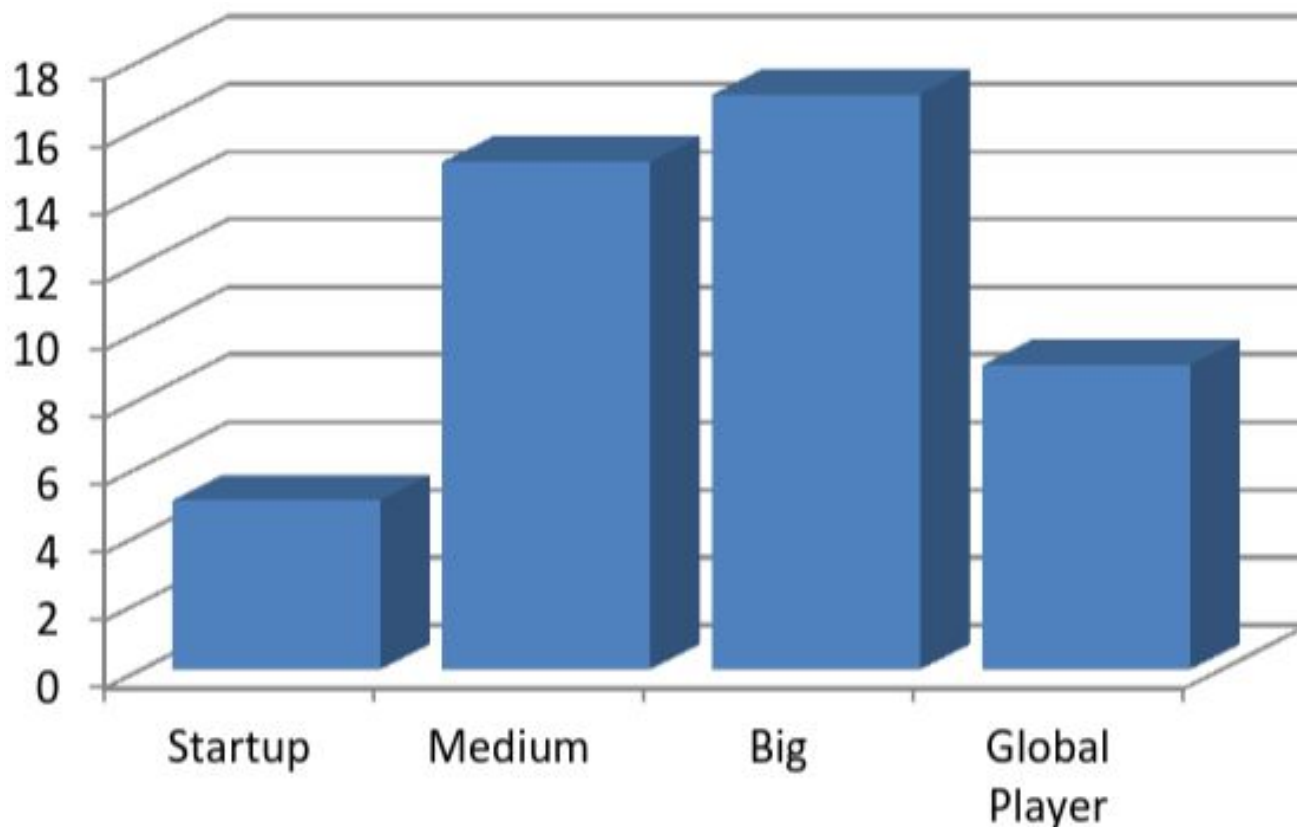
Revisitation: Size of vendors

- Despite startups and global players remaining the same,
- Changes suggest that the market is growing and companies are developing
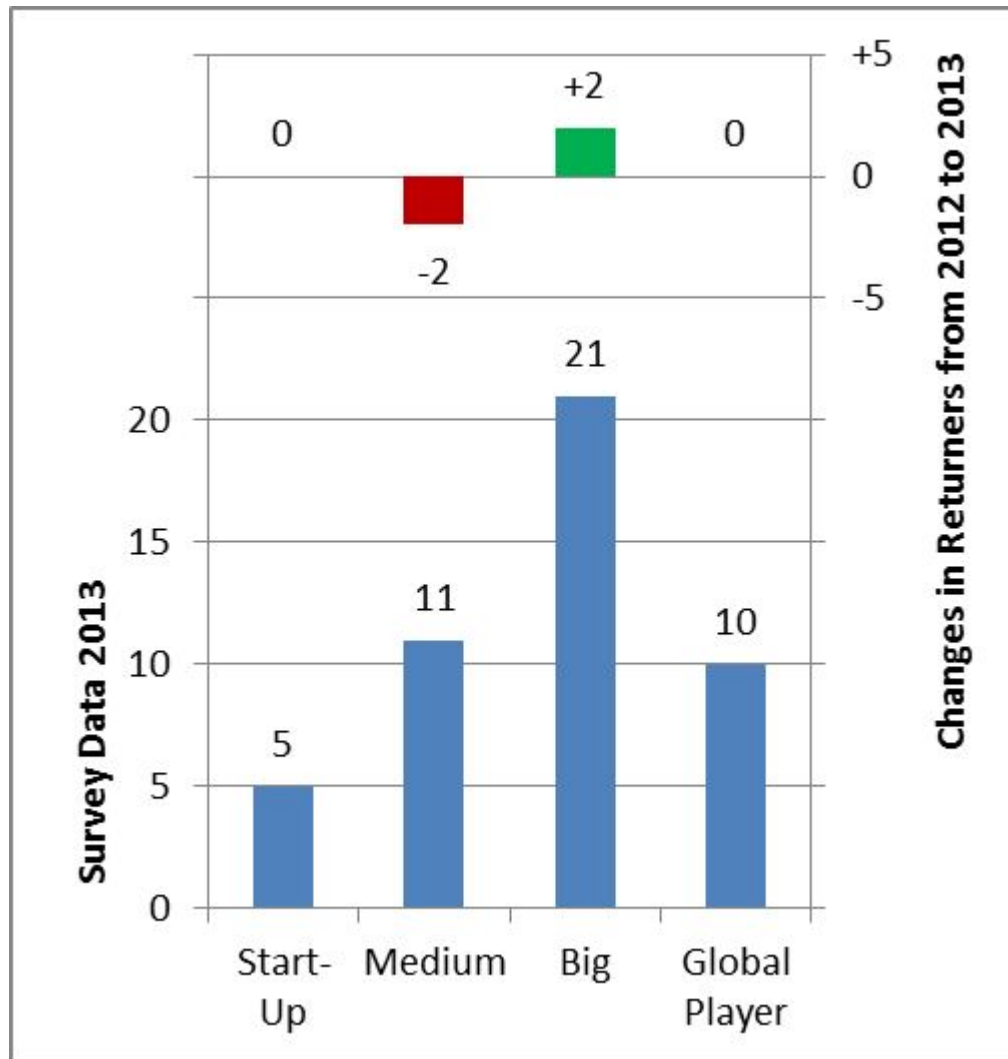
Findings: Maturity

Subject dimension about whenever level of establishment they have in the market
- A high maturity in projects signals that the space for new companies to enter the market is relatively small.

Revisitation: Maturity

- Maturity has not changed tremendously compared to 2012
- 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.
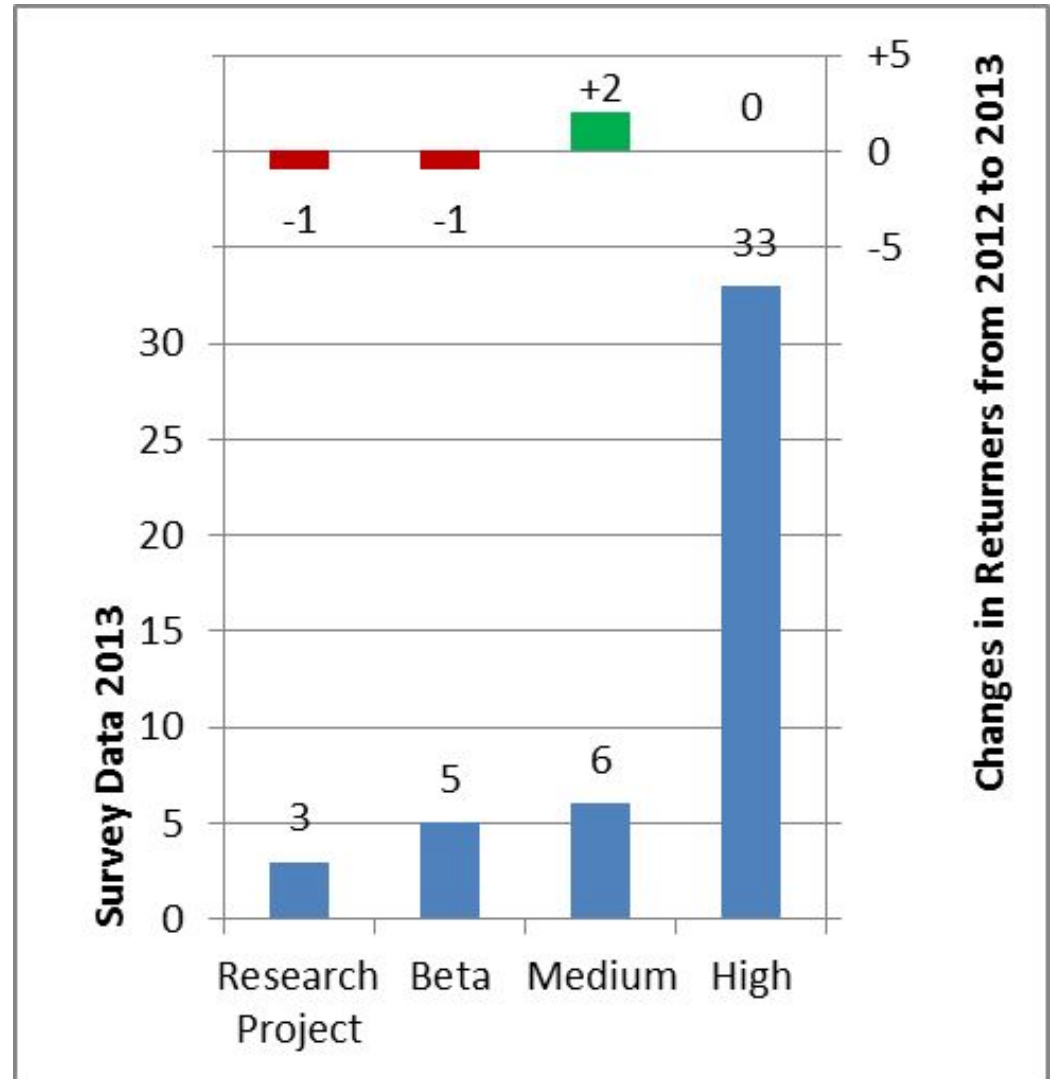
Conclusion

- The study presented a survey drawing a preliminary picture of the current data vendor marketplace
- It focused on the provider view of data marketplaces
- It may be interesting to study the market from the customer view and make distinctions between them too
- The 2013 revisitation of the paper gave us more insight on the market
- It analyzed trends and growth of the market by comparing the set dimensions a year later
- It shared a brighter light on the Data marketplace, and its current growing state in the mainstream market.

References

[1] Fabian Schomm, Florian Stahl, and Gottfried Vossen. 2013. Marketplaces for data: an initial survey. SIGMOD Rec. 42, 1 (March 2013), 15–26. https://doi.org/10.1145/2481528.2481532.

[2] Stahl, Florian, Schomm, Fabian and Vossen, Gottfried, (2014), The data marketplace survey revisited, No 18, ERCIS Working Papers, University of Münster, European Research Center for Information Systems (ERCIS), https://EconPapers.repec.org/RePEc:zbw:ercisw:18.