

An Unknown Treasure – How Do Companies Determine The Value Of Their Data?

BARBARA ENGELS

Researcher, MSc., German Economic Institute, Cologne

GERMANY

engels@iwkoeln.de

Abstract: The collection, analysis and exploitation of data are key drivers of the digital economy. But the importance of data for economic success is also increasing in industries that are not primarily associated with the digital economy. The basis of efficient data management is the evaluation of the data processed by the company. This study provides details of the data evaluation methods currently used by German companies. For an empirical analysis, a total of 1,235 firms from different sectors of manufacturing industry and industry-related services were surveyed using the IW-Zukunftspanel 2018. The study asks whether, and for what purposes, companies of different sizes and with different degrees of internationalization and digital orientation determine the value of their data. It also examines the characteristics which lead a company to prefer one data evaluation method over the others. Thus, an overview of the status quo in German corporate data evaluation is built up. The empirical analysis shows that data evaluation is still a side issue for German business. Most companies neither analyse their data now nor intend to do so in the future. Of the companies in this sample, it is mainly those that offer digital products that evaluate their data. Companies are currently unable to capture the potential of the data they possess and hence are unable to exploit the economic potential of digitalization.

Keywords: Digitalization, Data Governance, Big Data, Data Management, Data Economy

1. Introduction

The key drivers of digital transformation are the analysis and processing of data. In the digital economy, data is a central production factor. Even in sectors not primarily attributable to the digital economy, data collection, evaluation and analysis are becoming increasingly important (Yin/Kaynak, 2015). In addition to increasing efficiency and improving production, data management can enable new types of business models. Furthermore, data trading is becoming increasingly important (IDC, 2017). Different companies offer a huge amount of data, including address, market, consumer and spatial data (Dewenter/Lüth, 2018, 20). Data platforms from Industry 4.0, where machine-generated, non-personal data are very important, are less strongly represented so far. One reason for this is that many companies in these industries do not yet manage their data efficiently.

The basis for efficient data management is the ability of the company to evaluate its own data. It needs to be clearly determined what a data set contains and which value it holds for different stakeholders. If a company wants to use data as a production factor, it needs to be able to price its data in order to include it in the production as a cost or profit factor. An evaluation of the data is also indispensable for transferring data to other entities, for example in the context of data trade. Knowing how to determine the value of the data being used by the company contributes to the data sovereignty of the company, and thus to the ability to control data and to use it skillfully and effectively.

This study analyses the extent to which German companies from manufacturing industry and industry-related services sectors determine the value of the data generated or processed in their companies. Through logistic regression analyses, the factors that contribute to the ability and willingness of a company to measure the value of its data are determined. It is also analyzed which methods of data evaluation the companies use and which purposes they pursue through attaching a value to their data.

To the knowledge of the author, this is the first study that empirically demonstrates the status quo of data evaluation – and hence the potential for data analyses and digital transformation itself. The sample that this analysis is based on – small and medium-sized companies belonging to the manufacturing sector and related sectors – underlines the contribution of this paper, since these companies do not directly belong to the digital

economy but have a high digitalization potential within the framework of Industry 4.0. The “data treasure” these companies can take hold of includes data generated by sensors, machine running times and downtimes, and data on product characteristics. This data is used, for example, to make processes more efficient and effective. It can also foster new business models. The analysis and evaluation of these large data sets is becoming a critical success factor.

2. Data

The following analysis is based on an enterprise survey. The sample used is based on the IW-Zukunftspanel conducted in spring 2018. The IW-Zukunftspanel is a regularly conducted representative survey of managing directors of industrial and industry-related service companies in Germany.¹The survey includes general data on revenue, industry, age and management structure of the company as well as questions on internationalisation, research and development and innovation activity. In spring 2018, the companies were additionally surveyed on their data stocks and evaluation.

For the empirical analysis, a total of 1,235 firms from different branches of manufacturing industry and industry-related services were considered. The sample is not representative. Large companies are deliberately overrepresented in order to be able to make comprehensive and broadly-based statements about this subgroup as well. However, this distortion does not play a role in the regression analyses, because they are controlled regarding to the industry and size.

The range of sectors in which the companies surveyed operate is broad. A quarter of the companies surveyed belongs to the metal and electrical industry (excluding mechanical engineering). Almost one fifth belongs to business-related services such as information services, management consulting, auditing, research and development and marketing. As industry-related sectors, logistics and wholesale trade (14 percent) as well as media, information and communication technology (6 percent) are included.

More than two-thirds of the companies have less than 50 employees and one of three even has less than ten employees. One tenth has at least 250 employees. Most of the companies (88 percent) surveyed generated an annual revenue of less than 50 million euros. 27 percent had an annual revenue of up to 1 million euros. With a correlation coefficient of 0.73 there is a clear and statistically significant correlation between the number of employees and the revenue.

The sample also differentiates between digital and non-digital companies. It is assumed that there are different affinities for data evaluation depending on the degree of digital orientation, since data and data sovereignty are particularly important in the digital economy.² The survey included an evaluation of the company's own digital maturity as well as the extent to which digital products are part of the company's service offer. Both can, but do not have to be related. Particularly in Industry 4.0, companies focus on digital value-added processes, but often produce non-digital end products or end products with only a few digital components.

The digital maturity of companies was measured by the strength of digitalization of the processes and products (goods or services). 1,177 of the companies surveyed provided information on whether they have virtualised their products, processes or tools in full, in part or not and whether they work with digital models, or whether their business models are based on data models, data analyses or algorithms. 984, or 84 percent, of these 1,177 companies are not rated as digitalized, but as computerized. They therefore tend to use digitalization only partially and in a supportive way (they are rather digitized than digitalized). A clear minority of 16 percent or 193 companies is digitalized. In order to determine the degree of digitalization regarding the company's product offer, the share of total revenue that can be attributed to either entirely digital goods and services (e.g. software, data models, web design), to products with digital components, or to non-digital products is examined. Non-digital companies are those whose revenue is 100 percent attributable to non-digital products. More than one third of the companies surveyed belong to this group. Only 4 percent of the companies have a revenue that is almost entirely attributable to digital products. 11 percent have a “digital” revenue of more than 50 percent. Overall, the companies surveyed provide digital goods and services only to a moderate level.

There is no strong correlation between digital orientation and the degree of product digitalization in this sample. The correlation coefficient between the two variables is 0.31 and the relationship is statistically significant. Among digital companies, only 15 percent offer only non-digital products; among non-digital companies, the share amounts to 39 percent. All in all, the companies surveyed in Germany are classified as

¹ Lichtblau/Neligan, 2009.

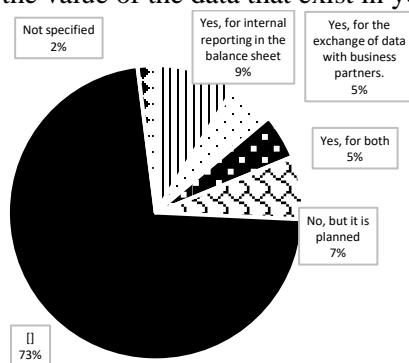
² BVDW, 2017 and IDC, 2017.

having a low to moderate digital degree. This also renders them very relevant for an analysis of the significance of data evaluation because particularly companies that are still at the beginning of their digitalization possess large data stocks generated in the production process that are neither analysed nor put into efficient use.

3. Data Evaluation Among Companies

A very broad understanding of data has been used in order to determine the extent to which companies determine the value of their data. In this analysis, “evaluation” means defining an economic value, i.e. a monetary value, based on various factors such as quality and currency. Data are per se very heterogeneous. Depending on the source, generator, storage type, analysis method or purpose, numerous classifications can be made:³ Data can be classified according to their structure (unstructured, structured or semi-structured), their format (e.g. text, image or video file), their reference (person-related, potentially person-related or not person-related) or its generator (machine-generated or not machine-generated). There is also a physical, semantic or syntactic understanding of data. In the present survey, it was left to the respondents' judgement how they view the question they were asked, "Do you determine the value of the data that exist in your company?". It remains unclear how raw or processed the data is that the respondents had in mind. A detailed query of the considered data sets would be desirable because the relevance of the data evaluation probably varies significantly depending on the data characteristics. However, this goes far beyond the scope of this analysis. The empirical analysis shows that the evaluation of data is anything but prevalent among the surveyed companies (Figure 1): Almost 80 percent of the companies do not determine the value of their data, and almost three quarters do not have any plans for a future evaluation either. The fifth of all companies that evaluates data has different evaluation purposes: 9 percent of these companies evaluate their data for internal accounting and reporting purposes, 5 percent for the exchange of data with business partners and 5 percent for both.

Figure 1: Data evaluation in companies: Share of companies that evaluates data or not (question: "Do you determine the value of the data that exist in your company?"); n=1,235



Logistic regression analysis is used to examine which company characteristics influence the probability that a company determines the value of its data (Table 1). The aim is to identify which companies are more likely to be pioneers in this area. The dependent variable is the binary variable "Evaluation of data yes/no". Three models are estimated:

- **Model I - General** contains the general evaluation as the dependent variable. The enterprise evaluates the data for internal reporting, for the exchange of data with business partners, or for both purposes;
- **Model II - Internal** contains the internal evaluation as the dependent variable. The company evaluates the data only for internal reporting;
- **Model III - External** contains the external evaluation as a dependent variable. The enterprise evaluates the data only for the exchange with business partners.

The control variables are the number of employees (different size categories), the revenue (different size categories), the sector, whether companies are engaged in research and development, whether they are innovators, whether they operate internationally, are digitally mature (see above), and what proportion of their

³ Dewenter/ Lüth, 2018, 5.

revenue is attributable to digital products (see above; different size categories). The category not listed in each case is the basic category to which the resulting estimates refer. This means, for example, that for the control variable "number of employees 10 to 49" the estimated value is compared to those referring to the number of employees in the not specifically listed basic category "less than 10 employees". The numerous control variables reduce the number of companies included in the estimation to 960, as some did not answer the relevant questions. A multicollinearity test showed no alarming multicollinearities that could distort the estimation results. Robustness tests with regressions in which only revenue or number of employees were controlled, but not both at the same time, also lead to reliable results. In order to interpret the results not only qualitatively but also quantitatively, odds ratios were established.

Table 1: Logistic regression result (odds ratios⁴).

***/**/* Significance at the 1-/5-/10 percent level; Standard error in brackets. Number of companies: 960.

	General (I)	Internal (II)	External (III)
Number of employees (basis: less than 10 employees)			
10 – 49	0,858 (0,245)	1,029 (0,331)	0,659 (0,242)
50 – 249	0,566* (0,188)	0,834 (0,309)	0,443* (0,189)
250 and more	1,655 (0,732)	2,432* (1,211)	1,132 (0,615)
Revenue (basis: small - up to 1 million Euro)			
Medium-sized (1 to 50 million euros)	2,051** (0,621)	1,452 (0,487)	1,926* (0,743)
Large (50 million euros and more)	1,291 (0,600)	0,528 (0,288)	2,270 (1,272)
Sector (basis: chemistry/pharmaceuticals)			
Metal/Electrical Industry	0,762 (0,298)	0,785 (0,332)	0,824 (0,492)
Other industry (excl. mining)	1,660 (0,714)	1,509 (0,705)	2,232 (1,401)
Construction	0,361** (0,185)	0,392* (0,218)	0,498 (0,370)
Logistics/Wholesale	1,148 (0,487)	1,230 (0,566)	1,841 (1,128)
Media/ICT	0,331* (0,188)	0,398 (0,252)	0,389 (0,311)
Industry-related services	0,726 (0,310)	0,526 (0,252)	1,110 (0,689)
Mechanical engineering	0,796 (0,374)	0,800 (0,407)	0,930 (0,649)
Research⁵	1,042 (0,245)	1,066 (0,279)	0,958 (0,301)
Development	0,927 (0,213)	1,114 (0,287)	0,702 (0,211)
Innovator⁶	0,842 (0,168)	0,828 (0,185)	1,090 (0,285)
Internationalisation⁷	1,050	0,889	1,323

⁴ Odds ratio is a measure of association in which two odds are compared with each other. Odds are quotients of the probability that an event will occur and the probability that it will not occur. Odds ratio for the evaluation of data shows by how much greater the probability is that a company will evaluate its data if that company meets a particular property (e.g. being digitally mature) compared to the group without that property.

⁵ Companies with continuous research or development had corresponding expenditures in the years 2015 to 2017. Where such expenditure was not identified every year, research/development was described as occasional.

⁶ Innovators are companies that have introduced new or significantly improved products, services or processes since 2015.

	(0,220)	(0,209)	(0,364)
Digital maturity	1,176	1,042	1,182
	(0,132)	(0,137)	(0,165)
Product digitalization⁸ (basis: non-digital)			
Digital > 95%	4,067**	3,015*	4,803**
	(2,234)	(1,923)	(3,510)
Digital 50 to 95%	6,390***	4,064***	10,590***
	(2,346)	(1,684)	(4,860)
Digital/digital components > 50 %	3,595***	2,774***	4,695***
	(0,964)	(0,821)	(1,716)
Digital/digital components 10 to	2,303***	2,061**	2,354**
	(0,606)	(0,592)	(0,887)
Digital/digital components < 10%	1,525	1,396	1,902
	(0,534)	(0,534)	(0,929)

Table 1 shows which factors significantly increase the probability that a company determines the value of its data. A value of 1 means that the probability is equal to the reference group. This analysis allows the following conclusions:

- The probability that a company evaluates its data increases significantly if the company offers a relatively large number of **digital goods and services (product digitalization)**. However, the strongest effect is not seen among companies with only digital products, but among those that generate more than 50 but less than 95 percent of their sales with digital products. The probability that a company generally evaluates its data increases by a factor of 4.1 if a company generates 95 percent or more of its sales with digital products, and by a factor of 6.4 if the revenue share of digital products is between 50 and 95 percent. This is in comparison to companies with 100 percent non-digital products. However, the small number of cases and the high number of standard errors must be taken into consideration. The probability that a company evaluates its data increases by a factor of 3.6 (2.3) when a company generates 50 percent or more (between 10 and 50 percent) of its revenue with digital and partially digital products. In addition, in the models that analyse internal or external data evaluation, these effects are statistically significant and strongly positive. The highest effects are achieved in Model III (external evaluation). The probability even increases 10-fold if a company owes 50 to 95 percent of its revenue to digital products.

- The **revenue** has a strong significant effect on the evaluation probability. Again, the highest category value does not produce the strongest effect (except for the internal evaluation). Companies with particularly high sales show no significant effects. On the other hand, the effect for medium-sized companies is significant compared to small companies. The probability that a company evaluates its data increases by a factor of 2.1 if the company has an average revenue (compared to a company with a low revenue of up to EUR 1 million). A similar effect results for the external evaluation.

- Also for the **number of employees**, there is no estimated effect that constantly increases with the category values. A mid-level category has the only significant value. Revenue and the number of employees therefore influence the probability that companies evaluate their data. However, this does not apply to all size categories. The probability that a company evaluates its data in general changes by a factor of 0.6 if the company has 50 to 249 employees; this means eventually a decrease in probability. The probability of the external evaluation (Model III) changes by a factor of 0.4, i.e. it also decreases relative to the reference group. The estimated value of the internal evaluation (Model II) is not significant. However, in Model II the probability of internal evaluation increases 2.4-fold if the company has 250 employees or more (with a high standard error).

- There are some significant differences depending on the **sector**, but these are much smaller than those for product digitalization and revenue. It is remarkable that companies from the media and ICT sectors, which are most likely to belong to the data-based digital economy, evaluate their data significantly less often than companies from the reference group chemistry/pharmaceuticals. The probability that a company evaluates its data only increases by a factor of 0.4 or 0.3 compared to the reference group if the company belongs to the construction sector or the media/ICT sector. For the construction industry, the probability of internal evaluation changes by only 0.4. Compared to the reference group the probability decreases.

⁷ Non-international companies are those with no foreign activity. Weakly internationalised companies are those with an export volume of less than 25 percent of turnover. Strongly international companies have an export volume of more than 25 percent of revenue and they have production or research and development partly abroad.

⁸ Product digitalization categories are formed according to the percentage of revenues generated by digital products.

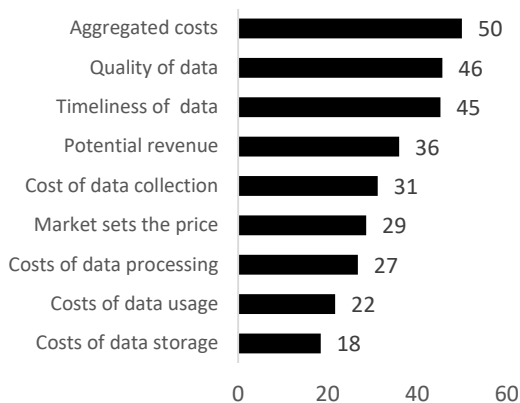
- Research and development, internationalization, being an innovator and being digital mature have no significant impact on the probability that a company evaluates data. The key driver of data evaluation is the degree of **product digitalization**. This is particularly relevant when focusing on the evaluation of data for the exchange with business partners (external evaluation). The fact that product digitalization has a considerable influence on the probability of data evaluation is logical: Digital products often include or require the analysis and evaluation of data. This applies either to the product itself (in the context of its use) or to the further development and optimisation of the product. If, in an alternative model, the revenue is omitted, for example in order to obtain a higher number of cases, a significant effect of **digital maturity** is obtained. The probability that a company will generally evaluate its data increases by 1.3 when it is digitally mature, compared to only computerized companies. This also applies to the external evaluation. The internal evaluation does not show a significant coefficient. The quality of the reduced model is lower than that of the initial model. Therefore, the more comprehensive model with the lower number of cases is used.

5. Methods of Data Evaluation

The companies that determine the value of their data were also asked which factors and methods they use to do this. These include cost-based assessments (by cost of collection, storage, processing, use), market-based assessments (potential turnover, market pricing) and property-based assessments (quality and timeliness of data). For simplification, the terms method, factor and aspect are used synonymously in the following. Overall, only a selection of methods is considered in this paper. A full analysis would exceed the scope of this analysis. Figure 2 shows the results based on the IW-Zukunftspanel.

Most companies (50 percent) evaluate their data according to the costs of collection, preparation, use and/or storage. It is reasonable to combine the different types of costs to aggregated costs, as it is difficult to clearly differentiate the costs. 46 percent of the surveyed companies evaluate their data according to quality, followed by timeliness (45 percent). In total, 50 percent of the companies evaluate according to the two related criteria of quality and timeliness. Revenue-oriented evaluation methods follow. Among the cost-oriented methods, the evaluation according to the costs of data collection is particularly relevant; more than 30 percent of companies evaluate data using this aspect. The costs of data storage are almost irrelevant for the companies surveyed.

Figure 2: Methods of Data Evaluation. Share of companies according to data evaluation method, in percent. Multiple answers possible. Number of companies: 218.



On average, the companies consider three factors or aspects when determining the value of their data. Only one fifth uses merely one evaluation method. 12 percent of the companies evaluate their data, but do not consider any of the above aspects. These results show that several aspects play a significant role in data evaluation. There are numerous data evaluation methods that this study could not take into consideration.

There is no difference in the choice of the evaluation method depending on whether the evaluation is made for internal or external purposes. Both the companies that evaluate their data for internal purposes and those that evaluate them for external purposes prefer an evaluation according to costs, followed by quality and the timeliness of the data. Also, among those who evaluate data both for internal and external purposes, these preferences prevail.

Table 2. Methods of evaluation according to company characteristics. Results of chi-squared tests on the equality of means of different subsamples (depending on whether company falls into the respective category or not).

***/**/* Significance at the 1-/5-/10 percent level.

	Costs	Data		Potential revenue	Market price
		Quality	Timeliness		
Digital maturity	0,151*	0,122	0,127	0,197**	– 0,027
Product digitalization	0,036	– 0,113	– 0,111	– 0,049	– 0,138
Research	0,159**	0,204****	0,185**	0,155**	– 0,128*
Development	0,144**	0,059	0,052	0,068	– 0,108
Innovator	0,180***	0,105	0,174**	0,126*	– 0,129*
International	0,130*	0,020	– 0,074	– 0,104	0,011
Many employees	– 0,025	0,128	0,168**	0,095	– 0,048
High revenue	0,054	0,112	0,174**	0,089	0,018

Table 2 contains the results of chi-squared tests on the equality of means of the response values of different subsamples of the surveyed companies. These tests can reveal differences between companies according to their characteristics. For this purpose, the surveyed company characteristics (apart from the sector) were defined as two categories per variable. A division of the variables into more than two categories would have led to very small subsamples in the mean value tests, as only a maximum of 218 enterprises commented on the methods of data evaluation. Therefore, regressions based on a small number of cases would not lead to reliable results and are hence dispensed at this point. Instead of distinguishing between four different employee categories, the companies were divided into those with up to 49 and those with at least 50 employees (“many employees”). This limit was also chosen in order to achieve a sufficiently large number of companies in the different categories. A distinction was also made between companies with either small revenue and companies with medium or large revenue (“high revenue”). Companies with digital products are those that generate at least 10 percent of their revenue with digital products. For example, row 1 of Table 2 shows to which extent a statistically significant difference exists when a company determines its data according to costs, data quality or timeliness, potential revenue or market price or not, depending on whether the company is digitally mature or not.

- **Digitally mature** companies evaluate their data significantly more often according to potential revenue than companies that are not digitally mature. They evaluate them mostly according to quality. This is not shown in Table 2, but in descriptive statistics that are not listed for the sake of space. The same applies to the preferred methods/criteria in the following sections.

- The fact that a company generates less or more than 10 percent of its revenue from **digital products** has no statistically significant effect on the popularity of the chosen data evaluation method. Companies with digital products most often evaluate their data according to the general costs.

- Companies with **research** activities evaluate their data less according to the price set by the market than companies that are not active in research. Research companies evaluate more according to the quality of the data, the actuality of the data, the costs in general, the potential revenue and the costs of the collection than non-research companies. For research companies, the quality of the data is most likely to determine the evaluation and not the market price.

- Whether a company is engaged in **development** or not contributes to small differences in data evaluation. There is only a significant difference in the costs, which developing companies choose significantly more frequently than non-developing companies. Developers most often evaluate according to the general costs.

- When a company is regarded as an **innovator**, it evaluates its data more in terms of general costs, timeliness of data, potential revenue and cost of use than a non-innovative company. Innovators evaluate most often according to the timeliness of the data.

- **Internationally** operating companies evaluate their data significantly more often according to the costs than nationally operating companies. They are most likely to evaluate their data based on general costs.

- Companies with a particularly large number of **employees** rate the timeliness of the data significantly more often than companies with few employees. It is their preferred method of data evaluation.

- The same applies to companies with **high revenues** compared to companies with lower revenues, except that companies with high revenues primarily evaluate the quality of the data.

Overall, there is no clear preference among companies for an evaluation method. This indicates how complex the evaluation of data sets is in practice. It is remarkable that innovative, researching and digitally mature companies evaluate the quality and timeliness of the data. In addition to the quality and timeliness of the data, the costs also matter. Cost evaluation is a more rational, conservative evaluation method, which is not necessarily effective for data trading, where the potential benefits of both the supplier and the customer, and thus the potential revenue, must be considered. However, the companies surveyed tend not to use market-related evaluation methods. Only digitally mature companies are pioneers in this area and evaluate their data significantly more often according to the potential revenue the data could generate.

6. Conclusion

This empirical analysis concludes that the evaluation of data among German industrial companies and companies from industry-related services sectors is still a marginal issue. Most companies do not evaluate their data and do not intend it either. Among the companies in this sample, it is mainly those that offer digital products that evaluate their data. Data that is collected along the value chain is still a black box for most companies in industry and industry-related sectors. Companies are currently unable to capture the potential of the data they possess. Only a few companies obviously know what their data or data sets are worth.⁹

If German industrial companies and companies from industry-related sectors evaluate their data, they usually use more than one method for data evaluation. A preference for a specific method is not evident from the analysis. This underlines the complexity of data evaluation in practice. Most of the times, companies evaluate their data either according to costs in general, or according to the quality and timeliness of the data. It is advisable to provide companies with standardized assessment tools in order to better assess the quality and timeliness of the data. Standards in data evaluation could also lead to more consistent evaluations and better comparability of data sets, which could simplify their trading and handling inside and outside the company. Organized data infrastructures that offer a clear technological and legal framework could, in addition to further research in this area, lead more companies to use the opportunities offered by data management and consequently better exploit the potential of digitalization.

⁹ Short/Todd, 2017.

References

- [1] BVDW – Bundesverband Digitale Wirtschaft. „BVDW-Umfrage: Datensouveränität von zentraler Bedeutung für Digitale Wirtschaft.“ Accessed October 23, 2018. <https://www.bvdw.org/der-bvdw/news/detail/artikel/bvdw-umfrage-datensouveraenitaet-von-zentraler-bedeutung-fuer-digitale-wirtschaft/>.
- [2] Dewenter, Ralf and Lüth, Hendrik, “Datenhandel und Plattformen, ABIDA – Assessing Big Data Gutachten 01IS15016A”, Accessed October 21, 2018. <http://www.abida.de/de/blog-item/gutachten-datenhandel-und-plattformen>.
- [3] IDC – International Data Corporation and Open Evidence. “European Data MarketSMART 2013/0063. Final Report”. Study for the European Commission (Directorate-General for Communications Networks, Content and Technology). Brussels, 2017.
- [4] Lichtblau, Karl and Neligan, Adriana (Ed.). „Das IW-Zukunftspanel. Ziele, Methoden, Themen und Ergebnisse“. Cologne, 2013.
- [5] Short, James E. and Todd, Steve. “What’s Your Data Worth?”, MIT Sloan Management Review, March 3, 2017, <https://sloanreview.mit.edu/article/whats-your-data-worth/>.
- [6] Yin, Shen and Kaynak, Okyay. “Big Data for Modern Industry: Challenges and Trends [Point of View]”, *Proceedings of the IEEE* 103, no. 2 (2017): 143–146.