# The Value of Data for the German Water- and Wastewater industry


**Lukas Hormann**

Hochschule-Ruhr-Werst, Duisburgerstraße 100, 45475 Mülheim an der Ruhr, lukas.hormann@hs-ruhrwest.de

**Sarah Stuhl**

Hochschule-Ruhr-Werst, Duisburgerstraße 100, 45475 Mülheim an der Ruhr, sarah.stuhl@hs-ruhrwest.de



*Abstract: This article deals with the value of data for the German water- and wastewater industry. It addresses questions and findings of the digitization to the special field of the economy of utilities and here for water- and wastewater. Firstly describing general aspects of how digital data is compiled and structured, the article shows the growing significance of digital data by analyzing the value chain of water- and wastewater companies and showing, where relevant sources of data are. After that, the assessment of the findings shows, data is becoming more and more important, although the German water- and wastewater companies are organized as a natural monopoly. As the use of (technical) digital data has a long tradition in the German water- and wastewater sector, the article gives suggestions on where new fields of usage of digital data can occur. The expansion of using digital data to meet challenges like cost-pressure, optimizing maintenance patterns and building new sources of value for the company are key findings. As the water and wastewater companies are at a starting point in digitizing the article comes to the conclusion that further research is necessary to foster the advanced use of digital data and give guidelines as well as information.*



*Keywords: Waste-water, Water, Digitization, Data, Value of Data, Water industry, Google, Facebook, Web-economy, Predicitve Maintenance, world-wide-web, Web 3.0.*


## 1 Introduction

The digitization is on everyone's lips and does not halt in renewing whole parts of web-based economies but also traditional industry sectors. How to conquer the upcoming challenges of digitizing in the German water- and wastewater industry is a very well discussed topic these days. For the reader sitting on the sidelines of the topic, this might be a surprise, as utilities are not expected to be the trailblazer of digitizing their businesses. However, the sheer amount of conferences, meetings, and conventions on the subject of digitization shows the ambition German Water- and Wastewater companies have internally as well as externally [1], [4], [5]. The outcome of any kind of activity in terms of digitization produces (digital-) data. However, digitizing the German water industry and following that, raising and storing digital data should not be at the end in itself. Therefore, the value of (digital-) data is a crucial topic that needs further discussion as the intelligent usage could, on one hand, lead to an increase of income (e.g. through new business models), on the other lead to the decrease of cost (e.g. through tapping further efficiency potential). It is clear, that web-based companies like Google or Facebook are using the data they raise, as it is part of their business model. The question this text will answer is how using digital data by sectors (e.g. the water industry) that were not web based in the first place, can be beneficial. Speaking in economic terms, the virtue of raising, storing and using of digital data is perceptible, when data has a value for the company. The question that arises; how can the value of data be measured and which kind of data is suitable to nurture the work of a water- or wastewater companies and what are the economic benefits of it?

For that reason, the following article will discuss the value of (digital-) data. It will give a brief overlook which type of data is common and how they evolved over the development of the "modern" world-wide-web.

# 2 The development of internet economy – from the first Browser to Web 3.0

The internet as we know it today was and is an ongoing development over the last 49 years. As the first network for computer dates back in 1969 developed for the US-Forces, it emerged in 1982 as the terms TCP/IP were introduced. Universities and other public utilities used the Internet (interconnected net) to share knowledge in the first place. The introduction of the WWW (World Wide Web) in 1992 and the invention of the web-browser (1993) were important steps in lifting the rudimental network to a new level. The commercial side of the internet was born [11]. From 1993 to 1997, the numbers of websites increased from about 130 to 1 million and the internet became a "new economy" itself. By introducing the new segment of the German stock market "NEMAX" in Germany in 1997 the elation about new business models and so-called "modern companies" found its realization. In the years from 2000 to 2002, the hype peaked in the DotCom bubble, which led to the closing of the NEMAX and a phase of consolidation. In 2005 a general recovery established the Web 2.0. The internet developed a new face with changed user- usability and perception.

Companies like Google, Facebook, etc. incorporated the user and created a web, which supported participation and an individual usage. Web Sites where no longer just a source of information, entertainment etc. but rather an interconnection between the user and the data behind the site. It was now possible to create own content (i.e. a post or comment) in blogs, wikis, podcasts, and communities and get an individual response. The possibility to individualize the content of web pages created the term “Web 2.0”.

In German, this phase of the internet is also known as the web of participation. Although the “Web 2.0” describes the internet to the current date, the next development (Web 3.0) will eradicate the weaknesses. Those are most notably the sheer amount of unsorted information (data) and its restricted possibility to search it and get an exact result. The more data or information is stored on the web, the less accurate an individual search for information can be. The solution for that is semantic search. It seeks to improve the search accuracy by understanding the searchers intent and the contextual meaning to generate more relevant results [3]. Nowadays some major web search engines like Google or Bing incorporate some elements of semantic search. However, the development is still ongoing [10]. To what extend semantic search will modify the value of data seems to be a wild guess. It can be expected, that the demand for personal data will increase (as the semantic search uses personal data in particular) and therefore its value. Nevertheless, before it is possible to assess the value of data for the German Water- and Wastewater industry, the term data should be analyzed and categorized.

# 3 The diversity of data

Digitization and interconnection are the basics for web-based economies. Digitization dematerializes and virtualizes processes in value chains, and thereby influences the velocity, productivity, flexibility and dispersion of economic activities. The interconnection represents the premise for an efficient working-method at a high level of work sharing and guarantees a facilitated access to information (data). The internet (the addition of digitization and interconnection) is, therefore, a global marketplace for information, which are the assets of web-based economies [10]. The digitization of the physical goods water- and wastewater is not possible. Delivering water and deriving wastewater are processes whereas a digitization is partly possible. The monitoring of a channel system with multiple sensors could be an example for that. If the importance of data in the new era of digitization is that high, a general description of what data is and how categorizing it can help becomes evident.

As information and data are frequently used as equally understood terms, research about the definition of information and data comes to different interpretations. A general definition we first want to introduce is a rather technical; Witt (2010) says, "Data are statements composed of interpreted signs resp. signals without any context" [2]. He states that data become information when they "are interpreted in a contextual way and lead to an increase of knowledge (esp. if it happens as a process).

With this wide definition of the term "data", we would like to address the categorization. How can we differentiate various datasets?

The following Figure 1 categorizes different types of data.

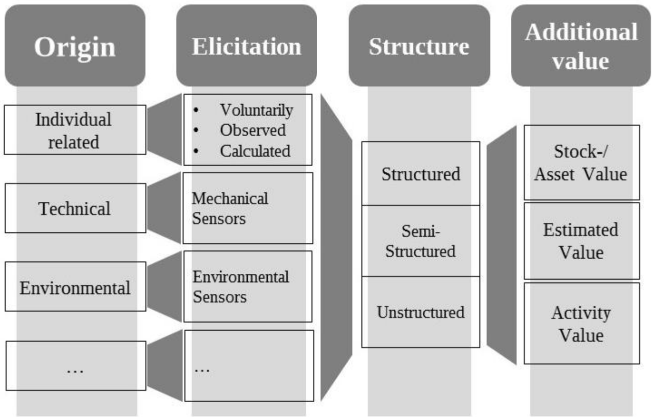


Figure 1
Categorizing data [Own Illustration]

In giving, a more general orientation in which forms data can be structured, the interpretation of the image follows. Every kind of data has an origin. It comes from either a person, a technical or an environmental input (imagining other sources is possible as well). Where data is ascertained there has to be a source for it. The categorization by source makes it easier to anticipate for what data can be used and how the quality is.

The next category elicitation is how the data is compiled. Especially for individual related data, three types of elicitation are common [12]. Data collected on a voluntary level happens e.g. when a social media profile such as Facebook is set up. The collection is directly and mostly with a benefit for the person revealing the data (participating on the social-media-platform). Two rather indirect forms of collecting data are the possibilities of observation and calculation. Observing the browser-activity via cookies or the activities on the website generate data about interests and activities of the user [9]. The use of a fitness tracker or an application that produces positioning data etc. could be an example as well. The third option how personal data is compiled is a calculation out of several different datasets.

This could be a combination of voluntarily or observed data as well as third-party data like information given by other users etc. An example could be the combination of the age and gender out of a social-media-profile and the viewed products on an e-commerce platform. This leads to a new, more exact dataset.

For the structure of data sets the three technical categories structured, semi-structured and no structures are common. Fasel and Meier are defining them as follows:

- Unstructured data: no structure at all or no general format
  - Video
  - Audio
- semi-structured: information in a defined order but not suitable for the (end-)user
  - E-Mail
- Structured data: defined length, defined format (numbers or words in a certain length a so-called “string”)
  - E.g. the date in DD/MM/YYYY [6].

About 20% of the worldwide digital data is structured and stored in relational databases (esp. SQL-Databases). However, semi- and unstructured data are stored in NoSQL-Databases, which break with the tradition of relational databases although relational systems are still the most common. Relational databases lack performance when data is used intensively. Examples of that kind of applications are websites with heavy data loads as well as streaming-media-applications [6]. The increasing number of media applications on the web can deduce expectations of a rising number of NoSQL-Databases in the next years.

The last part of the categorization aims to show different ways evaluating data. Answering the question how the (economic) value of data is measured is not easy and has a great span of interpretations. The context in which data is compiled, measured and used, affects the value and makes the evaluation a complex task. Figure 1 shows the chosen categories strategic asset, estimated value and activity value. Data as a strategic asset means the monetarization of e.g. customer information. This is a well-practiced action especially in the area of advertisement and direct marketing. Generating new information by combining existing data can also lead to a gain in value. The estimated value is a clear speculation of the upcoming value certain datasets can gain in the future. The last category “Activity Value” means the benefit data can generate when it is used more than once. The costs for compiling, storing and managing data are high whereas the marginal cost of using is nearly zero. Therefore the multiple uses of data or even the ambition to do so (e.g. geo-data for a navigation application) can make it more valuable [11].

# 3 Datasets of the water- and wastewater industry

To draw the line back to the water and wastewater-industry we first have to address which type of data the industry is gathering before thinking about evaluation. To put things in a broader perspective, it is useful to understand the value chain and its particular parts. Analyzing each step allows a more in-depth view where data is or could be gathered and in a further step, what value can be anticipated. Figure 2 shows an overview of the value chain and its parts and gives examples for sources of data.

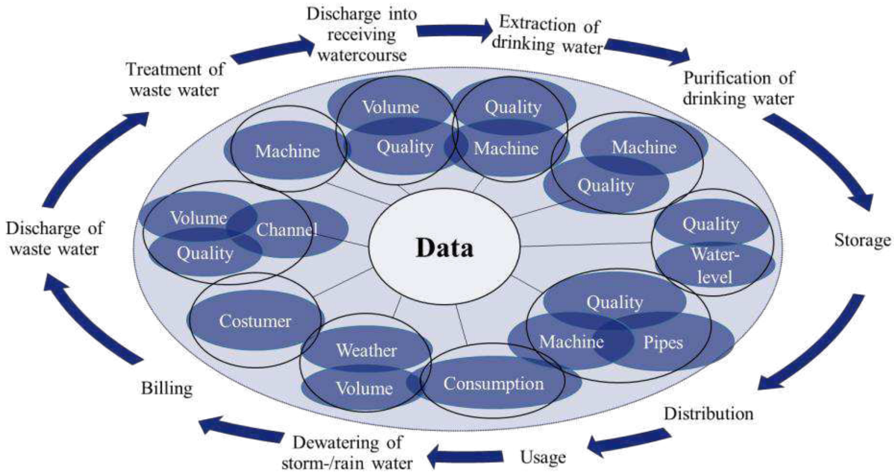


Figure 2
Value Chain of the water- and wastewater industry [Own Illustraition]

- Extraction of drinking water:

As seen above, the value chain can be started at the step of extraction resp. production. Freshwater can have many different sources (groundwater, surface water, etc.) therefore the monitoring of the quality and machinery is important. The quality of the extracted water determines which effort has to be taken to purify the water for further use in the processing. The data collected can be used to predict maintenance issues and achieve an overall improvement of maintenance patterns.

- Purification of drinking water:

The processing of the freshwater takes place in a water plant and uses various techniques to remove unwanted substances. Important parts are the sterilization and disinfection. Therefore, the observation of the produced water quality as well as the used machinery produces various types of (technical) data.

- Storage:

Following the processing, the next step is (when needed) the storage of water. Mainly for maintaining a certain level of pressure in the water distribution system, it is also needed when an unexpectedly big amount of water is demanded. The reservoirs and its sensors can produce quality-data as well as water level-data, which again are rather technical aspects.

- Distribution:

The distribution of the freshwater follows the next step. The observation of the pipe-system, the quality and the pumps for the distribution can be seen as further sources of data. Especially the monitoring of the distribution system could produce (if applied area covering) big amounts of datasets. The case that a distribution system is fully equipped with sensors is relatively rare but seems to be a topic in the future. Analyzing the data of the pipe-system could be used to optimize the management of maintenance in the first and lead to an automated maintenance-system in the second place.

- Usage

After delivering the freshwater to the customers, they use it in many different ways. One interest could be about the different consumption patterns, especially for industrial customers. Again the equipment with sensors of every customer is rarely or even unknown to this date.

- Dewatering of storm- or rainwater:

As dewatering refers to the drainage water from sealed surfaces, data of the volume of stormwater and therefore weather data could be interesting to predict loads of rainwater.

- Billing:

Especially in the billing process, the use of customer data is obvious. It is clear that information like bank details, etc. are kept in a database to facilitate the organization. Keeping in mind, that depending on how many customers are connected, the event of billing becomes more complex. The connection of customer data with e.g. consumption patterns can lead to new insights.

- Discharge of wastewater

Measuring the volume and quality of discharged wastewater is uncommon and therefore not a well-used source of data. Some utilities use area covering sensors in their sewage system to establish a real-time-controlled channel system. In cases of heavy loads, the intelligent control can reallocate volumes of wastewater to take the load of intensively used parts of the channel system. Therefore such a system could provide a great amount of data as well.

- Treatment of wastewater / Discharge into receiving watercourse

After discharging and collecting the wastewater, it has to be treated in a purification plant before another discharge into a receiving watercourse is possible. Similar to the treatment of fresh water, the treatment of sewage is a complex, stepwise process. The used machinery can produce different kinds of technical data which helps to monitor and control the processes of the treatment. As the last step, the discharge into a receiving watercourse takes place. The measurement of the quality of the effluent and its volume are datasets that may be produced.

To put things in perspective, the majority of produced or potentially produced data is technical and appears in the competence of the companies. It is collected by sensors measuring quality, operation condition and volume. The elicitation of customer data is relatively rare at this juncture. Nevertheless, the introduction of smart-water-meter might change that in the near future [9].

# 4 The value of data

As shown datasets can occur in every step of the water cycle, the question that arises is how they have a value or even add a value for the companies. Therefore the next chapter deals with how the described sorts of data can be beneficial.

When thinking about the digitization some of the first things that come to mind are new ways of businesses like Facebook, Google, Amazon, Uber, etc. All these business models base upon the interaction of people with digital data and the sharing of information either between users among themselves or with the provider of a service that happens digitally in the first place. By reserving a ride with Uber or placing an order on Amazon, the user does that via a digital channel and when the service is bad, the user has the possibility to find alternatives. The water- and wastewater industry, however, have few incentives to digitalize its business. On the one hand, this is justified by the fact that water itself cannot be dematerialized (as mentioned above) on the other the fact that the water industry is a natural monopoly, and so customers do not have a choice to swap the service provider, may foster the lack of innovation.

Despite this finding, the industry has a long tradition of using data for its benefit. The focus is on the utilization of technical data for gaining efficiency and therefore reducing costs in the service of maintaining to purify and distribute water or collect and treat wastewater.

The presented data inputs “quality”, “machine” and “volume” consequently are the one with the longest tradition and at the time given the most promising to gain value in using technology coming from digitization. Terms like Big-Data, predictive maintenance, artificial intelligence, etc. come to mind when thinking about that. With that background, the assumption that the value of "technical" data will remain high and increase further becomes clearer. This fits the categorization data as a stock- or asset value. The combination of data (even if not possible today) can create new knowledge that helps to improve current tasks and resolve old issues. Especially the topic of cost pressure developed over the last years and seems to be an overall concern of water- and wastewater utilities [7].

Cost pressure could be caused by several reasons externally as internally. For instance, a company might see itself in the obligation to increase the pension provisions because of the constant low level of interest rates. The expected profit might be used for that which leads to fewer earnings for the municipality and therefore for the call to decrease costs. Another consideration could be the fact that most companies deal with decreasing demands for drinking water hence decreasing amounts of wastewater which leads to profit cuts. The digitization can help to optimize the efficiency and effectiveness of processes. The mentioned optimization of process costs might be a solution for that. In addition, the development of new business models has to be taken into account as well. On one hand, the analysis of or even the possibility to purchase new customer-data can help to anticipate new necessities on the other might there be the possibility to monetarize current datasets when a demand from other industries occurs (e.g. web-based economies).

For web-based economies, personal data is the fuel for their business models [8]. The way the water- and wastewater industry can address this topic lies more in the dark than the described technical aspect. Although individual data is kept and dealt with (e.g. in the billing process) further applications are missing. An area covering rollout of smart water meter could lead to individual consumption analysis which then again could lead to new models in businesses. To put oneself in the perspective of the costumer and reveal its necessities is the task the companies have to face. Undoubtedly the intelligent use of the data the companies already have and the ones that they will compile in the future have the opportunity to be a real added value not only for the industry itself but for the customers and the society.

**Conclusion**

Ever since the development of the modern internet and web-economies, digital data plays a significant role in every kind of business or private use. The determination of its value whereas is a topic which comes to surface since data can be analyzed in a more efficient and significant way. While the digitization produced various new forms of businesses and creation of value, the water- and wastewater industry, however, is at a starting point of exploring "advanced" solutions of digitization. This is not only because of the reason how the industry is organized but also how data can and is ascertained The number of sources along the value-chain shows, that there are possible ways to gain further information about the own business and the customers. The efficient and intelligent use of the raised data is the assignment the water and wastewater industry has to work out. By achieving a new way of dealing with digital data the possibilities of enhancing the added value for the own business, the customers and the society as a whole.

**References**


[1] Annual Conference on Water Economics, BDEW, Berlin, Germany, 24th and 25th of September 2018, https://www.bdew.de/media/documents/20180608_WWJT-Programm-2018.pdf, last downloaded on 22nd of June 2018.

[2] Bernhard C. Witt: Datenschutz kompakt und verständlich. Eine praxisorientierte Einführung, Wiesbaden, Germany, 2010, pp. 4-6.

[3] Bettina Fazzinga, Thomas Lukasiewicz: Semantic search on the Web, In: Semantic Web-Interoperability, Usability, Applicability, vol. 1, Amsterdam, The Netherlands, 2010, pp. 1-7.

[4] Conference on Industry 4.0, VKU, Berlin, Germany, 15th of February 2018, https://www.vku.de/verband/veranstaltungen/industrie-40-konferenz-2018/, last downloaded on 22nd of June 2018.

[5] Conference on Water Economics – Digitization, IWW, Mülheim at the Ruhr, Germany, 1st of March 2018, https://muelheimer-tagung.de/, last downloaded on 22nd of June 2018.

[6] Daniel Fasel, Andreas Meier: Big Data. Grundlagen, Systeme und Nutzungspotenziale, Wiesbaden, Germany, 2016, pp. 3-35.

[7] Erik Gawel, The Regulation Framework of the German Water Sector from an Economic Point of View – Taking Stock and Perspectives, In: gwf-Wasser | Abwasser, vol. 5, Essen, Germany, 2016, pp. 538-556.

[8] Laurence Morvan, Data: The fuel of the Digital Economy and SME Growth, Paris, France, 2016, pp. 1-4.

[9] Mark Oelmann, et al.: Smart Water, In: energie | wasser-praxis, vol. 4, Bonn, Germany, 2018, pp. 38-43.

[10] Rodica Tirtea, Claude Castelluccia: Bittersweet cookies. Some security and privacy considerations, Heraklion, Greece, 2011. pp. 3-13.

[11] Stefan Gutberlet, Alexander T. Nicolai, Heinke Röbken: Determinanten der Martkeintrittserfolge von imitativen Startups. Eine empirische Untersuchung über die Geschwindigkeit bei der Unternehmensentstehung und ihr Einfluss auf den Unternehmenserfolg, Wiesbaden, Germany, 2011, pp. 58-61.

[12] Wolrd Economic Forum: Personal Data: The Emergence of a New Asset Class, Cologny/Geneva, Switzerland, 2010, pp. 6-7.