Why Do We Need Data Privacy?

Volker Klingspor(✉)

Bochum University of Applied Sciences, 44801 Bochum, Germany
volker.klingspor@hs-bochum.de
http://www.hochschule-bochum.de/fbw/personen/klingspor.html

Abstract. In recent years, various socio-political debates and scandals have raised old and new questions regarding data protection that, among other things, will also lead to new European legislation initiatives. However relevant each issue may be, there is far too little discussion in the public which potentials, be it positive or negative, exist with the possibility of combining data from different sources. In this article I want to give a non-exhaustive overview of the manner in which such information about everyone of us is collected today, before I discuss the social risks this may entail. I close the article with some theses outlining a path that helps to protect the rights of freedom of the citizens despite the extensive collection and analysis of data (My heartfelt thanks goes to Edward Sodmann for proofreading this text. He required tons of hours to generate something from my text, that can be understood at all.).

Keyword: Data privacy

1 Current Trends

Information processing is increasingly integrated into everyday consumer electronics. While the average Internet user is worried about his data collected by Amazon or posted on Facebook, information about him is being collected and interpreted extensively in much more private and sensitive areas. Examples are:

1.1 Communication Data

WhatsApp, Twitter, and Facebook not only save the data we post about ourselves, be it the already quite personal daily joggs or photos of their loved ones. Hardly anyone is aware that a personality profile with information about their marital status (marriage, children, etc.), education, consumer behavior, etc. can be determined solely from the use of language, without actually communicating this information [19]. So we unconsciously give many more details about us than we realize at the time of communication.

Email is a highly decentralized form of electronic communication. There are many providers or one can even host an email server oneself. Thus, there is no easy way to access all emails of a person. It requires a relatively deep intrusion

S. Michaelis et al. (Eds.): Morik Festschrift, LNAI 9580, pp. 85–95, 2016.
DOI: 10.1007/978-3-319-41706-6_3

into the network infrastructure (as the NSA has shown us) to analyze email communications on a large scale. Moreover, it is very easy to create different email accounts for different purposes. This makes it even more difficult to create a comprehensive profile of an email user.

The conventional forums allow for a certain anonymity. Forums are typically themed: I am interested in motorcycles and diving, so I can log into two completely independent forums using different pseudonyms. But again, it is of course possible to use the IP address to link the various pseudonyms to the same person.

The communication via social networks and communication platforms such as Facebook, Twitter and WhatsApp has influenced the ability to collect data dramatically. Social networks thrive on the fact that they are, on the one hand, people-oriented, and, on the other hand, forming communication monopolies. Only if most of the communication is done via a single platform, the platform makes sense. Facebook, Twitter and WhatsApp are successful because *almost all* use the same platforms – similar platforms such as the VZ platforms have been forced out of the market. However, the fewer platforms exist, the more they know about us.

No one wants to have to post the same information on different platforms just to reach different friends. Then one could just use email. But this monopoly pushes towards a centralization of data. More so since the purchase of WhatsApp by Facebook, and (after a cooling-off period) the subsequent pooling of data from the two platforms, Facebook knows a great many details of billions of people: their lives, their preferences, their behavior, their whereabouts, and, in particular, their social networks.

A Facebook user profile can draw fairly accurate conclusions about the sexual orientation, ancestry, religion, political attitude, personality, intelligence, well-being, age, gender, relationship status, and the drug abuse of the user. All this information can be obtained indirectly, without asking the user any questions whatsover [19]. With Facebook knowing twenty percent of the world population, this is no trifle matter!

1.2 Photos and Videos

While public video surveillance is subject to strict conditions and hence the data protection is essentially ensured, countless pictures and videos are posted on all kinds of communication platforms in the so-called private sector. Furthermore, ubiquitous smartphone photography and filming with action cams is steadily increasing. Each pastime is now captured on video without considering the people that may unknowingly be in the background. The fact that, unlike photography and classic filmmaking, is especially critical concerning action cams because no conscious image design takes place. While image design usually tries to have as little distraction in the picture as possible, action cams record everything that happens to be in front of the lens.

This is particularly problematic since we have little impact on the publication of pictures in which we happen to be coincidentally. While we can ignore

Facebook and the like (albeit considered as old-fashioned or nutters), we can only object to the publication of photos if we become aware of a photo, and if we are more than a random accessory in the background. While this used to be no problem, as long as the photos were developed on paper and only shown in the circle of friends, digital images quite often appear on the Internet.

Photos and videos are no highly critical information as such. Linking names to photos and videos leads to a critical conflict situation in the near future. The more photos are tagged, the better the face recognition method of Google, Facebook, etc. will perform. In turn, images that have not yet been tagged, can be assigned to specific people. Similarly, as the search for buildings with a photo is available on Google, it is a likely scenario that in a few years one can find the names of the people who are on a photo or video. In addition, it will be feasible to find pictures or movies with non-public persons on the net – possibly only because they appear randomly somewhere in the background of a picture or video.

Since the photos and videos are often marked with metadata such as date and location, more information, such as motion profiles, can be extracted or extrapolated.

1.3 Smartphones

In 2014, worldwide about 1.85 billion people use smartphones [16]; in Germany there are approximately 41 million people [15]. Apps on these devices are easily able to collect very detailed profiles about their users – without them knowing it. A look at the permissions of the first six matches for Android flashlight apps reveals that these can all access the Internet, and five of the six apps require additional permissions. The app with the most permissions, for example, can read the device status and device ID, change system settings, retrieve running applications, take pictures and videos, as well as read, modify, and delete arbitrary files on the device.

For an app that is supposed to only turn on the LED, it's a lot of rights! This begs the suspicion that this app is to produce not only light in the darkness of smartphone owners, but also light in the dark of the supplier by consistently uploading information about smartphone usage so that the supplier is able to analyze the behavior of the user. Ironically, this app is rated above average, so that it is installed probably more often than other, perhaps less curious apps.

Unfortunately, the user of an Android device is not empowered to withdraw special rights of an installed app (unless he roots his smartphone). Apple with their iOS is far ahead: once an app wants access to resources of the smartphone, the user is asked in advance.

1.4 Internet of Things

More and more everyday items are based on electronic control modules. Thus, for efficient energy use one's own home becomes a networked, intercommunicating system. Devices will cooperate in the future in order e.g. to adapt the current electricity consumption to the power generation status or to weather conditions. Heaters will automatically adjust to the habits of the inhabitants.

And of course, these devices must always communicate with the owners. Either explicitly through concrete instructions ("the washing should be done by 16:30"), or by recording the behavior ("if it is recognized that no one is at home on Tuesdays, the heater control can respond accordingly"). Most times, increased comfort is made possible by the extensive collection of data. If one can control blinds and light via a smart home functionality via the Internet, a lot can be recorded detailing the course of our days.

Again, of course: as long as the data remain stored locally, and do not leave their homes, this is not problematic for the time being. However, it is already being planned for the detection of current consumption that information about the load behavior in households are stored centrally in order to facilitate the control of power plants and grids. Fortunately, this issue is still in the hands of the legislature, so anonymization and aggregation of data are regulated [7]. In the consumer sector, however, the authorities cannot respond to each new product with new legislation. And here a technology invades our households whose potential goes far beyond the Orwellian fantasies. Today's television sets can be controlled by gestures or voice commands. This is especially concerning when the TV can be switched on via gestures or language because this requires that a camera or microphone is continuously in receiving mode. TV sets recognize who sits in front of them and report the appropriate user to the Internet services so that the viewer can use these services without prior log-in. At least the manufacturer Samsung stipulates that in the terms that the recorded data may be sent to third parties [9].

Another example of the increasing networking of everyday objects is the rising number of equipment in vehicles with communication modules. Today, all modern cars record not only functional errors but also parameters of driving behavior. High-end vehicles are fitted with communication modules that are able to transmit these data online to the manufacturer. Together with the driver recognition via electronic key or seating positions information can be generated from the speakerphone's address data or from the GPS data that go far beyond mere motion profiles. Without knowing a concrete study, I am sure that the mood of the driver can be recognized by the accelerator and brake protocol. In contrast to the opportunities that arise here, the possibilities of black boxes for detection of driver behavior as they are offered by commercial truck insurers are less threatening.

1.5 Linking Information

To date, information substantially exists as data islands. Every company, every forum, every game portal, each social media site collects data about customers or users. For some of these providers, we may mask our identity with pseudonyms. Once we enter into a business transaction, this is generally not possible. In addition, Facebook in particular expects and reviews the application with a real name, although this is actually not necessary. Research from 2006 shows that in an individual case it is possible to relate different, even anonymous user profiles with each other, matching it to a specific person. The best-known example is the assignment of a IMBD profile to a video rental customer's account [13]. While in this example, only a single person was de-anonymized, whole subnets of the Flickr network could be de-anonymized in [12]. In 2000, Latanya Sweeney showed that 87 % of all Americans could be uniquely identified using only three bits of information: ZIP code, birth date, and sex [17]. One year before, she identified and related the anonymized medical records of the governor of Massachusetts [6].

Meanwhile various approaches to de-anonymize people are developed by combining distributed data. Researchers can given credit that they want to demonstrate to users the possibilities and to suggest a less generous attitude with their data. Nevertheless, it is expected that within a few years, techniques will be on the market that link user and search profiles on a large scale. Information which is regarded as confidential and anonymous by Internet users can be assigned to them anyway after all.

1.6 Pre-crime Detection

Data Mining techniques are now being used for the prevention of crime and terror. "The Future Attribute Screening Technology project (FAST) system has the capability to monitor physiological and behavioral cues without contact. That means capturing data like the heart rate and steadiness of gaze of passengers about to board a plane. The cues are then run through algorithms in real-time to compute the probability that an individual is planning to commit a crime" [10].

In Memphis, Tennessee (USA), data analysis has been used to preventively to monitor locations where potential offenses may be committed since 2005. In 2010, the police and IBM celebrated with a decline of serious crime by more than 30 %, including a 15 % reduction in violent crimes since 2006 [11]. The reason for the decline is, according to the publication, that particular gang disputes could be detected early. Unfortunately, it is not clear from text to what extent people were arrested preventively, and, conversely, to what extent honest persons preemptively avoid areas where they can potentially be arrested.

2 Why Is Privacy Important?

2.1 Loss of Autonomy and Freedom Rights

"I have nothing to hide!" This sentence is heard often at discussions about data protection and in particular on data that is to be given to law enforcement authorities. I would like to emphatically disagree.

"Privacy describes the extent to which a person other people are permitted to enter one's own world" [4,18]. We constantly negotiate the limits of our privacy and our voluntary disclosure with others but also with ourselves. The scope of one's privacy is very individual, and may even vary at different times. We need an open mental sphere to evolve. Young people need privacy to learn to think for themselves and to act self-confidently. The less a teenager has trust in his privacy, the less he will dare to act contrary to the norms of his peer group, and the more uniform and less confident he will be. But adults need privacy to personally and professionally develop, too. Sabine Trepte formulated as a benefit of privacy among others the autonomy to break social norms and to experiment with new behaviors and thoughts [18].

The German Federal Constitutional Court ruled in its landmark ruling on the census [1] that if someone is not reasonably certain how his personal information is used and shared, he can be inhibited in its freedom, to plan and decide in a self-determined way. People who do not know what information is held on them, will try to behave as inconspicuously as possible. There is the likelihood that central fundamental rights are waived without being aware.

The protection of privacy not only means that "thoughts are free", but also that I may share my thoughts in a safe room with the people that I trust. For this reason, the inviolability of the home is an essential and fundamental right. Equally important is the inviolability of communication. Of course, not every conversation, every chat, every posting is equally confident, nor any confidentiality is of equal importance. But I must be able in a figurative sense "to close the door" at all times in order to monitor accurately who participates in a conversation.

In academic literature, scholars distinguish three dimensions of privacy [14]: *Informational privacy* refers to the fact that my data will not be public unless I want to. *Decisional privacy* describes the right to be protected in decisions and actions from unwanted external influences. *Local privacy* describes the protection against the entry of other into private rooms and areas.

And only those that respect the privacy of a person, respect him as an autonomous person in the sense that he has the freedom to live his life independently and to seek his own happiness [14]. Conversely, secret knowledge about other people gives institutions power over these, which can lead to changes in behavior and behavioral adaptations. In this sense, to understand the ruling of the Constitutional Court: this power through secret knowledge threatens the freedom of expression and freedom of assembly, and thus central fundamental rights.

Unfortunately, many Internet users already resigned: *"It is already too late, they already know everything about me"*. Many people who do not agree with the data collection use this as an excuse and keep using the new communications media. I agree with that in so far as all information that we have disclosed via social media services, etc., cannot be taken back, i.e. cannot be deleted. This certainly does not mean that we should continue feeling unconcerned. The more information that is available about us, the easier it will be in the future to associate this information with other strings of information, and finally to generate complex personality profiles.

The sooner we begin to be careful with our data, the better. The human quality to forget or to put into perspective what is said over time is just an important tool for personal development, such as the aforementioned privacy.

Only trusting the fact that details about my actions and statements over time disappear from the memories, allows me to freely discuss. Unfortunately, the Internet and data collection companies do not forget what we have posted!

The safest option for confidentiality is the boycott of as many Internet-based services as possible. This way less or no information will be disclosed, of course, and cannot be misused. And although I certainly advocate a more restrictive use of the Internet: this path sooner or later leads to an exclusion from the peer groups that communicate via social-media platforms. If I'm not on Facebook, I do not get the information provided exclusively via this platform. Also note that not even the complete Internet abstinence really solves the problem; this strategy will only delay the collection of data and cannot really prevent it, either. Data are now collected at every traditional purchase, every contract, and any travel booking.

To illustrate this more vividly: you search old classmates for a class reunion? Check with a credit agency such as Bürgel or Creditreform. They know the complete move history, including any change of name of any bank client. Just give as reason for the request "business contacts", and – a couple of euros poorer – you know the current address of your old class mate and, in addition, you can verify if he is able to pay his bill himself? [8].

2.2 Transition from a State of Law to a State of Prevention

A constitutional state is different from a preventive state in that the former does not preemptively act against potential offenders – only the charge of an offence may lead to a penalty. Preventive actions always lead to a restriction of the freedom of action – especially for innocent citizens.

Not without reason will the establishment of any public video surveillance be discussed intensively – particularly from the people that have nothing to hide. There is a very fine line between security and upholding fundamental rights. In a constitutional state we are not able to prevent all crimes, nor will we be able to solve every crime because of the given legal means. However, the rulings of the Federal Constitutional Court (e.g. concerning data retention) seem to suggest that this tightrope walk works quite well – at least in Germany.

2.3 Desolidarization Society

Machine Learning techniques as well as traditional statistical techniques are often used for risk assessment. However, automatic risk assessment can cause certain members of the population to be unreasonably excluded from a fair treatment that is based on facts and not extrapolation. If scoring functions are used already, they should produce accurate results in every case. The well-known example of someone receiving no credit because he lives in the wrong area indicate the poor scoring functions. However, I do not understand why trained and experienced bank branch managers no longer have the power to overrule the scoring result. The individual applicant may have sufficiently good proof that ultimately speaks for getting his credit.

In the insurance industry a more accurate risk assessment leads to a loss of solidarity. The more closely I can determine the risk of insurance, and the more accurately I can set the premium on this risk, the more the idea of solidarity insurance is lost. In almost superfluous insurances such as the legal protection insurance (for "normal" individuals) this might be a trifle.

When, however, the risk assessment excludes participation from society, we lose a fundamental principle of our state. Consider, for example, the car insurance sector: the fact that the premium depends on the type of vehicle is unproblematic – cars that carry lower premiums can be an option. Since the current premium determination has started considering whether the parents also own a car, and how old the driver is, the premiums for liability insurance can easily vary from case to case by a factor of three to four. I can even drive as prudently and cautiously as I want to – I may be charged higher premiums just because the statistics speak against me and I may not be able to afford a car. A corresponding analysis in the field of health insurances might point at an even worse scenario for a nation.

In my view, we are losing the basic idea of an insurance. Their idea is to spread the risk as much as possible, to give everyone the same opportunities. If, due to individual misconduct, premiums rise modestly, that is certainly acceptable. But to exclude people from insurance policies putting up high premiums without letting them have an influence on the design of the premium does not correspond to my idea of our society.

3 Is Data Collection and Analysis to Demonize Generally?

Data analysis is just a technique, and is not bad in itself. In many situations, including those that will perhaps be viewed by the population as questionable data analysis carries overall positive aspects.

The quality of Google search results bases on a very detailed analysis of user behavior. The fact that almost always a relevant hit is already on the first results page, reflects the quality of this analysis. The extrapolation quality of the on-screen advertising is very high, too. There is, of course, the self-interest of Google because the ads usually must be paid only if they are clicked. You do not get anything for nothing – without collecting and analyzing data Google search works just not as well (the older readers among us remember surely still with horror at the previous attempts to adjust the keywords as closely as possible to the search target, to aid search engines to find something useful).

And I love the assistance provided by many vendors displaying similar products. I can browse through music, and come across new, hitherto unknown pieces. Invoices are archived for years by the manufacturers and may be accessed by the customer at any time. There is no need for him to file away the bills at home in folder. As long as the data is saved by the providers, there is little danger from them. Scary is the fact that much more data is collected about customers than is required for the sale. Amazon e.g. gathers intensive data about the reading habits of the users of its e-book reader [3].

And also in the scoring of credits, there are positive aspects of a systematic data analysis. The denial of a loan is not only to protect the financial lender against losses, but also very much the borrower against over-indebtedness. The more accurate the scoring process the better I can grant credits to those that do not overburden themselves with this loan. It is a pity, however, if the scoring algorithm is not able to justify the classification, and if, at the same time, the employees of the lenders are completely exempted of the freedom to lend based on personal assessments and counter-scoring results.

There are many other data analysis applications from which citizens benefit more or less directly. The analysis of traffic flows may optimize traffic control and planning, so that road users get faster and more energy-efficiently to the destinations. The analysis of the movement patterns of people in shops or at events such as conferences can lead to improvements in the placement of objects (sale items, break counters, ...).

In this sense, there is certainly an unlimited number of useful applications of data analysis that can help to cope with the challenges that we have to face because of limited and increasingly costly resources. But how should society deal with the conflict between benefits and risks?

4 Conclusion

Big Data cannot be stopped, and in many areas data analysis is really helpful. It also seems to point out that with the larger amounts of data the statistical errors become less relevant – that are so used to analyze more data, the better the results are.

However, we need an intense social discourse about the ways our data is handled. We need empowered and educated citizens who are aware of the dangers and consequences of data collection. Citizens who actually read the privacy statements of companies or the permissions of apps, and discard applications when they have doubts about the legality of the use. Citizens who understand the principles and apply encryption techniques, and quit service providers who do not offer any encoding. Citizens who consider carefully what information they publish about themselves. At the same time, we need citizens who critically evaluate information gathered from the Internet and that rely on different sources of information, and that are always aware of the existence of fake identities and propaganda dressed up as information.

For these citizens, however, we need a better infrastructure to ensure their data are safe. Easy to use encryption programs that allow local data and data stored in cloud storage to be automatically protected by encryption. Communication systems, including in particular email, encrypting messages end-to-end, and thereby allowing a simple but faithful key exchange. Although this software already exists, it is still underused, which is, in my view, due to complicated use for IT laymen. It would be desirable if pictures, videos and other documents could be provided with an expiry date, after which they can no longer be seen. This would ensure that missteps in the identification process, particularly of young people would not permanent part of their public life.

It is absolutely certain that we also need new laws. The Federal Constitutional Court took very clear position in 1983 with the census ruling ensuring the right to informational self-determination. All European countries are now following the European Data Protection Directive, published in 1994 by having adopted country-specific data protection laws. All these laws do no longer work effectively in the global world, as they are binding only for those providers who are headquartered in the EU. Those who do not want to keep to European law, need only to open their businesses in Tonga (or the US), and can then handle data in an unrestricted way. Furthermore, the penalties for offenses against the privacy law are too low to deter potential perpetrators. Since 2012, a new European data protection regulation has been discussed, in particular addressing the two points mentioned above [2]. First, the privacy regulation is applicable as soon as *product* is offered in Europe, regardless of where this occurs. Then, European citizens are protected against suppliers from countries that have weak privacy rights. On the other hand, a significant increase in the potential penalties is planned so that this right can be enforced.

We need to ensure that people will not be constantly put into a box. The larger the amount of data that can be used to analyze, the better purely statistical correlations work, and the less analyses that are based on models are used [5]. Since we lack the explanations for automatically-made decisions, a bank employee is unable to overrule the scoring decision because he cannot identify an inconsistency between decision and request. I believe we should have the right to oppose purely statistical assessment procedures that force the seller to disclose the reasons for a refusal.

Still less can it be permissible to criminalize people on the basis of statistical analyses. In a constitutional state an offender still must have actually committed a crime before he is indicted. Each statistics-based crime prevention that determines and monitors potential criminals leads away from the rule of law towards a prevention and surveillance state.

And finally, we need more learning techniques (and other techniques), which ensure a very early real anonymization of the data collected. Research by e.g. the Fraunhofer Institute for Intelligent Analysis and Information Systems shows, that an analysis of pedestrian flows is possible just on the basis of aggregated data that cannot not be de-anonymized [20].

References

1. BVerfGE 65, 1; Az. 1 BvR 209, 269, 362, 420, 440, 484/83 (1983)
2. Proposal for a regulation of the european parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (general data protection regulation) (2012). http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52012PC0011&from=EN

3. Alter, A.: Your E-Book is reading you. Wall Street J. (2012). http://www.wsj.com/articles/SB10001424052702304870304577490950051438304
4. Altman, I.: Privace: a conceptual analysis. Environ. Behav. **8**(1), 7–29 (1976)
5. Anderson, C.: The end of theory: the data deluge makes the scientific method obsolete. Wired **16**(7) (2008)
6. Anderson, N.: “Anonymized” data really isn’t - and here’s why not. Ars Technica (2009). http://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/
7. Bräuchle, T.: Die datenschutzrechtliche Einwilligung in Smart Metering Systemen - Kollisionslagen zwischen Datenschutz- und Energiewirtschaftsrecht. In: Pödereder, E., Grunske, L., Schneider, E., Ull, D. (eds.) INFORMATIK 2014: Big Data - Komplexität meistern, pp. 515–526. Lecture Notes in Informatics, Gesellschaft für Informatik (2014)
8. Dambeck, H.: Verschollene Mitschäler: Bärgel und Creditreform retten Klassentreffen. Spiegel online (2006). http://www.spiegel.de/netzwelt/web/verschollene-mitschueler-buergel-und-creditreform-retten-klassentreffen-a-452447.html
9. Eikenberg, R.: Spion im Wohnzimmer - Privacy und Sicherheit bei Internet-fähigen TVs. c’t (4) (2014)
10. Erickson, M.: Pre-crime detection system now being tested in the U.S. (2012). http://bigthink.com/think-tank/pre-crime-detection-system-now-being-tested-in-the-us
11. IBM: Memphis police department reduces crime rates with IBM predictive analytics software (2010). http://www-03.ibm.com/press/us/en/pressrelease/32169.wss
12. Narayanan, A., Shi, E., Rubinstein, B.I.P.: Link prediction by de-anonymization: How we won the kaggle social network challenge. CoRR abs/1102.4374 (2011). http://arxiv.org/abs/1102.4374
13. Narayanan, A., Shmatikov, V.: How to break anonymity of the netflix prize dataset. CoRR abs/cs/0610105 (2006). http://arxiv.org/abs/cs/0610105
14. Rössler, B.: Der Wert des Privaten. suhrkamp taschenbuch (2001)
15. Statista: Anzahl der Smartphone-Nutzer in Deutschland in den Jahren 2009 bis 2014. http://de.statista.com/statistik/daten/studie/198959/umfrage/anzahl-der-smartphonenutzer-in-deutschland-seit-2010
16. Statista: Prognose zum weltweiten Bestand an Smartphones von 2008 bis 2017. http://de.statista.com/statistik/daten/studie/312258/umfrage/weltweiter-bestand-an-smartphones/
17. Sweeney, L.: Simple demographics often identify people uniquely. In: Data privacy working paper. Carnegie Mellon University (2000)
18. Trepte, S.: Privatsphäre aus psychologischer Sicht. In: Schmidt, J.H., Weichert, T. (eds.) Datenschutz - Grundlagen, Entwicklungen und Kontroversen, pp. 59–66. Bundeszentrale für politische Bildung (2012)
19. Tufekci, Z.: Ein Datensatz mit X. The European (3) (2013). http://www.theeuropean.de/zeynep-tufekci/7065-gefahren-von-big-data
20. Wrobel, S.: Big Data Analytics - Vom Maschinellen Lernen zur Data Science. In: Plödereder, E., Grunske, L., Schneider, E., Ull, D. (eds.) INFORMATIK 2014: Big Data - Komplexität meistern, p. 53. Lecture Notes in Informatics, Gesellschaft für Informatik (2014)