TITLE 8
The importance of data quality for Big Data

Thomas HOEREN and Barbara KOLANY-RAISER*

Introduction

Yves Poullet is an internationally highly respected legal scholar, whose research in the field of data protection has chiefly investigated the unpredictable and rapid development of information technology and its influence on Privacy. We were privileged to collaborate with Yves Poullet on the ECLIP-Project and other research projects and appreciate him for his visionary and inspiring work. Therefore we feel very honored to be asked to write an essay in this book. With this essay on the topic of data protection, we want to express our gratitude for the pioneer work he has done with his research in data protection and information technology law.

Knowledge is power and the use of Big Data technologies promises new insights by finding correlations that can be used for various benefits in mountains of data. The potential that lies within certain information comes to light in the overall context of available data. The real value of data lies in its usability, because the quantity of data says nothing about its quality. To put it bluntly, Big Data can make the need for a hypothesis used to get an understanding of the relation between data and causation redundant. Big Data analyses have the ability to find and classify even insignificant patterns in large, unstructured data sets. But if the data used for analyses is inaccurate, the result can only give a trend and not provide an exact result. There is thus a tendency to make decisions based on correlations and not on real causations.

The use of Big Data technologies allows us to gain a new understanding by using these improved techniques for real-time analyses. If even big companies like Google, with its huge amount of data, can only achieve

* The authors would like to thank Tristan Radtke for his help in preparing this essay. Dr. Thomas Hoeren is professor at the Law Faculty and Director of the Institute for Information, Telecommunication and Media Law – Civil Law Department, University of Münster; Dr. Barbara Kolany-Raiser is scientific assistant at the Institute for Information, Telecommunication and Media Law – Civil Law Department, University of Münster

poor results with their Big Data analyses as the well-known example of Google Flu Trends has shown in 2013,[1] the data economy, i.e. the data processing industry has a reasonable interest in processing data with high quality. That means that the data must be correct and not outdated. When we think of medical Big Data and its invaluable use for purposes such as public health practice, biomedical research, improvement efforts or institutions' quality assessment, the shortcomings of data quality due to such sources as data error, record fragmentation, lack of standardization or missing information can have disastrous effects.[2]

Hence data quality is very important, and legal obligations fragments designed to ensure high data quality can be also found in security-relevant areas such as statistical authorities[3] or financial service providers.[4] Data convey information that in turn represent particular parts of reality. Therefore, factual accuracy requires the correct representation in the data set. Data quality is also an issue when you buy data or buy the results of a Big Data analysis. For the buyer it is essential to know if the data or the analysis results have the contractually stipulated quality. Besides, poor data quality can also have great effects on people who are affected by assessment results, as in the case of scoring, which may lead to unfair discrimination.

In the following this essay first provides an overview of the principle of data quality in relation to German Civil Law, especially the liability for inaccurate data. Regarding the general sanctions for the use of false data, the principle of data quality is important in German civil law. In particular, general civil law may give rise to liability, both regarding tort liability (sections 823, 824 *Bürgerliches Gesetzbuch* (BGB)) and contractual liability (pre-contractual diligence obligations under section 280 BGB).

Subsequently it is shown how the principle of data quality has been developed from its roots in US legislation 1974 in the still valid 'Privacy Act'.

Regarding data quality on European level the centers of attention are the 1995 EU Data Protection Directive (Directive 95/46/EC), which determines that "Member States shall provide that personal data must

---

[1] D. BUTLER, "When Google got flu wrong", *Nature*, 2013, p. 155.

[2] Sh. HOFFMANN, "Medical Big Data and Big Data Quality Problems", *Connecticut Insurance Law Journal*, 2014, p. 289 http://dx.doi.org/10.2139/ssrn.2464299 (accessed 15. Dec 2017).

[3] Art. 12 of Regulation (EC) No. 223/2009 of 11 Mar 2009, OJ L 87, pp. 169 et seq.

[4] Section 17 Solvency Ordinance of 14 Dec 2006, *Federal Law Gazette I* pp. 2926 et seq. and section 4 of the Insurance Reporting Ordinance of 18 Apr 2016, *Federal Law Gazette I* pp. 793 et seq.

be [...] accurate [...]" in Art. 6 (1) (d), and the General Data Protection Regulation (Regulation (EU) 2016/679 (GDPR)). The latter was issued in May 2016 and will be applicable on 25th of May 2018. It will replace the EU Data Protection Directive and is a major step towards the harmonization of data protection principles within the European Union. It contains the principle of data quality in Art. 5 (1) (d) GDPR, using the term 'data accuracy'. Moreover the essay will raise the question of whether it is the right approach to embed the principle of data quality in data protection law. To assess this question the essay will eventually investigate the relation between data quality and data protection.

CHAPTER 1. Data quality in German Civil Law

Traditionally, the only way of selling information was to sell books. However, the buyer's expectations of a book's content were only irrelevant requests for information[5] and not worthy of legal protection[6]. Exceptions could only be made in the case of a large number of printing errors, missing pages or completely obsolete statutes.[7] Alternatively, one worked with assurances and guarantees[8] or an independent consultancy agreement.[9]

The origin of this attitude can be found in Article 5 (3) clause 1 of the *Grundgesetz* (GG) [German Constitution], which privileges both the author and the book's publisher.[10] Hence, printing errors "can indeed be largely avoided by a customary and commercially generally acceptable method of production, although not with certainty. In individual cases,

---

[5] German Federal Supreme Court, Case No. VIII ZR 232/56, Vol. 4 1958 *NJW* 138ff, Judgment of 26 Nov 1957.

[6] H. P. Westermann in: Fr. J. Säcker, R. Rixecker (eds), *Münchener Kommentar zum Bürgerlichen Gesetzbuch*, München, C.H. Beck, 7th ed., 2015, p. 434, para. 73.

[7] Local Court Stuttgart (Germany), Case No. 11 C 6932/94, Vol. 9 1995 *NJW-RR* 565ff, Judgment of 12 Jul 1994; see also Fl. Faust in: H. G. Bamberger, H. Roth *et al.* (eds), *Beck'scher Online-Kommentar BGB*, München, C.H. Beck, 43th ed., 2017, p. 434, para. 70; U. Foerste, "Die Produkthaftung für Druckwerke", *NJW*, 1991, p. 1436; U. Huber in: Th. Soergel *et al.* (eds), *Bürgerliches Gesetzbuch im Einführungsgesetz und Nebengesetzen*, Stuttgart/Berlin/Köln, Kohlhammer, 12th ed., 1991, p. 459, para. 344.

[8] German Federal Supreme Court, Case No. VIII ZR 137/71, Vol. 19 1973 *NJW* 843ff, Judgment of 14 Mar 1973; criticized in Huber, *ibid.*, para. 19.

[9] German Federal Supreme Court, Case No. VIII ZR 20/77, Vol. 20 1978 *NJW* 997ff, Jugdement of 8 Feb 1978; cf. for more details J. Köndgen, "Die Haftung von Börseninformationsdiensten", *JZ*, 1978, p. 389; Chr. von Hertzberg, "Die Haftung von Börseninformationsdiensten", *Fachmedien Recht und Wirtschaft*, Heidelberg 1978.

[10] See criticism in Foerste, *ibid.*, pp. 1433 et seq.

therefore, it may be that trade and communication does not and may not rely on the absence of a single such error".[11] However, this cannot be applied to data providers in the age of Big Data, because, nowadays, data itself is becoming the subject of contracts.

According to the conception of the legislature, the traditional German sales law provisions (sections 433ff *Bürgerliches Gesetzbuch* (BGB) [German Civil Code]) only included the purchase of physical objects, but not the purchase of rights and other immaterial objects. Notwithstanding, since the law of obligations reform, data are now 'other objects' as laid down in section 453 (1) (second alternative) BGB, with the result that the normal rules on the sale of goods are correspondingly applicable to data.[12] Additionally, one can apply the German law of service contracts to data contracts, but someone who is selling data is more likely to be sued under German sales of goods law than under the law of services.

Furthermore, the rules of tort liability are obsolete too. The main German tort law rule, section 823 (1) BGB, is applicable only if one of the legally protected rights is infringed. These rights are life, body, health, freedom and property in the understanding of German law (i.e. physical objects) or an absolute right in the above-mentioned sense. Consequently, under section 823 (1) BGB, there is no claim for mere pecuniary loss. Yet, data as such cannot be seen as 'another right' because it is not characterized as an absolute right, which applies to everyone and provides the owner with the authority to use it as he likes. Thus, section 823 (1) BGB merely protects against the complete loss of data as a loss of property, because the loss of property does not require damage to an object itself, but any negative influence on the owners' desire to use that property.[13]

A different approach for protection of data and its quality can be seen in section 823 (2) BGB. It provides protection in case of the infringement of e.g. the *Strafgesetzbuch* (StGB) [German Penal Code]. In contrast to section 823 (1) BGB it can be the basis for damages for pecuniary loss. In the

---

[11] German Federal Supreme Court, Case No. VI ZR 223/68, Vol. 441970 *NJW* 1963ff, Judgment of 7 Jul 1970.

[12] RegE, BT-Drs 14/6040 24; M. STIEPER in: J. VON STAUDINGER (ed.), *Kommentar zum Bürgerlichen Gesetzbuch: Staudinger BGB – Buch 1: Allgemeiner Teil§§ 90-124 und §§ 130-133*, Berlin, Sellier/de Gruyter, rev edn, 2017, p. 90, para. 17; R. M. BECKMANN in: J. VON STAUDINGER (ed.), *Kommentar zum Bürgerlichen Gesetzbuch: Staudinger BGB – Buch 2: Recht der Schuldverhältnisse §§ 433-480 (Kaufrecht)*, Berlin, Sellier/de Gruyter, rev. edn., 2014, p. 453, para. 37; also, Düsseldorf Higher Regional Court (Germany), Case No 17U 167/09, 2010 BeckRS 09514, Judgment of 17 Feb 2010; Munich Regional Court I (Germany), Case No. 16 HK O 10382/08, 2009 BeckRS 88429, Judgment of 10 Dec 2008.

[13] Karlsruhe Higher Regional Court (Germany), Case No. 3 U 15/95, 1996 CR 352, Judgment of 7 Nov 1995.

view of data quality especially section 303a StGB can be relevant. It prohibits the deletion and suppression of data, but, and that is the significant limitation, only against intentional acts.

Section 824 (1) BGB protects the individual against the threat to a credit of a person or a company. These threats must result out of untruthful factual claims. Section 824 (1) BGB offers no protection against expressions of opinion and value judgments. However, the Federal Court of Justice states, that the assessment of raw data, e.g. in the case of scoring, is the creation and communication of value judgments.[14] Claims for financial losses as a result of reliance on advice can therefore usually not arise in the Big Data sector. According to the Court, an exception can only be made "if during the expression, in the recipient's view the elements of the opinion fade into background in the face of the underlying facts" (paragraph 11). Section 824 (1) BGB does e.g. not offer protection against the wrong factual basis of a decision, but against a scoring result refusing a financial credit to a company.

Likewise, the law concerning the right to carry on an established business in the sense of section 823 (1) BGB is usually not infringed, since Article 5 (1) GG "does not prohibit the dissemination of true and objective information on the market, which can be important for the competitive behavior of market participants, even if the content has an adverse effect on individual competitive positions."[15] These antiquated guidelines show that the German law of tort is worthless, too.

As far as European law is concerned, the European Data Protection Directive contains the first fragmentary legal safeguards of data quality. If one looks at German law, one finds that a proper implementation of the EU Data Protection Directive –contrary to other states such as Austria and the United Kingdom– has not happened.[16] However, the Federal Data Protection Act (*Bundesdatenschutzgesetz* (BDSG)) deals in approaches with the question of data quality. In section 28b BDSG the law, under certain circumstances, allows data to be used to collect or apply a probability value for a given future behavior of the person concerned. Although the regulation refers in particular to credit scoring, other applications are possible. The protection of data quality is ensured by the fact that the used data must be demonstrably significant for the calculation of the probability. In

[14] German Federal Supreme Court, Case No VI ZR 120/10, Vol. 30 2011 *NJW* 2204, Judgment of 22 Feb 2011; see criticism in Th. WEICHERT, "Scoring in Zeiten von Big Data", *ZRP*, 2014, pp. 170 et seq.

[15] *Ibid.* (German Federal Supreme Court).

[16] Section 6 of the Federal Law on the Protection of Personal Data (*Federal Law Gazette I* No. 165/ 1999).

this respect the burden of proof lies in the extent of the Big Data evaluator. Beyond that, not only can a breach be fined (section 43 BDSG), but also lead to civil liability (section 280 (1), 241 (22) BGB).[17]

CHAPTER 2. Data quality in international regulations

The historical roots of the data quality principle have its origin in US legislation. In 1974 the still valid 'Privacy Act'[18] entered into force. It only contains legal provisions regarding data processing by federal executive branch agencies. Hence data processing by private sector parties does not fall within its scope. The US Privacy Act establishes certain requirements regarding the processing of personal data, such as "accuracy, relevance, timeliness and completeness as is reasonably necessary to assure fairness".[19]

With the Data Quality Act (DQA) or Information Quality Act (IQA), the Congress wanted to ensure that the information disseminated by federal agencies is accurate. Therefore the Office of Information and Regulatory Affairs (OIRA) within the Office of Management and Budget (OMB) was established to oversee the adoption of information quality guidelines by federal agencies. The quality guidelines are intended to guarantee the quality, utility, objectivity and integrity of information disseminated by federal agencies and to provide mechanisms for affected persons to correct inaccurate information.[20] The basic standards of quality consist of utility, objectivity and integrity:

- 'Utility' refers to the usefulness of the information to its intended users, including the public.[21]
- 'Objectivity' involves two distinct elements, presentation and substance. Objectivity includes whether disseminated information is being presented in an accurate, clear, complete, and unbiased manner. In

[17] M. KAMP in: H. A. WOLFF, St. BRINK (eds), *Beck'scher Online-Kommentar Datenschutzrecht*, BDSG, München, C.H. Beck, 22th ed., 2016, par. 28a ref. 180-181.

[18] http://www.archives.gov/about/laws/privacy-act-1974.html (accessed 15 Dec 2017).

[19] 5 U.S.C. 552 a (e) (5) concerning the processing of data by state 'agencies'.

[20] White House, Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies, https://www.whitehouse.gov/omb/fedreg_final_information_quality_guidelines/ (accessed 15 Dec 2017).

[21] https://www.whitehouse.gov/omb/fedreg_reproducible (accessed 15 Dec 2017).

addition, objectivity involves a focus on ensuring accurate, reliable, and unbiased information.[22]

- 'Integrity' refers to the security of information to prevent information from being compromised through corruption or falsification.[23]

The DQA applies only if the communication is intended for the public. Hence communication internal to the agency is excluded from the scope of application.[24]

In the US, data protection law is not regulated in one comprehensive federal law but has federal and state laws as well as guidelines. These contain provisions regarding data quality, such as the Fair Credit Reporting Act or the Health Insurance Portability and Accountability Act of 1996.

The OECD has long recognized the need for international data protection regulation and adopted the above-mentioned US principles and extended them in the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data in 1980. Although the guidelines are not binding under international law, they have contributed in an essential way to establish data protection as an object of international regulations.

While the DQA in the US regulates quality standards only for public sector information, the OECD Guidelines are applicable to personal data, whether in the public or private sectors. Paragraph 8 established the data quality principle which names important elements like accuracy, completeness and contemporaneity. The Guidelines have been reviewed and updated in 2013.[25] In 1980 the OECD also adopted a second recommendation concerning the "15 principles on the protection of personal data processed in the framework of police and judicial cooperation in criminal matters",[26] which contains in its principle no. 5 an even more detailed description of data quality. This principle does not only establish that reasonable steps should be taken to ensure that personal data is accurate and up-to-date, but data must also be evaluated. In case of inaccuracy or incompleteness the principle determines that the data concerned must be erased or rectified.

---

[22] https://www.whitehouse.gov/omb/fedreg_reproducible (accessed 15 Dec 2017).

[23] US Department of State, Information Quality Guidelines, https://www.state.gov/misc/13864.htm (accessed 15 Dec 2017).

[24] For more detailed information about the background of DQA and thereby caused changes of agency actions: Al. N. HECHT, "Administrative Process in an Information Age: The Transformation of Agency Action under the Data Quality Act", *J. Legis.*, 2005, p. 233.

[25] OECD, Privacy Guidelines of 2013, https://www.oecd.org/sti/ieconomy/privacy-guidelines.htm (accessed 15 Dec 2017).

[26] OECD, 15 Principles on the protection of personal data processed in the framework of police and judicial cooperation in criminal matters http://www.statewatch.org/news/2007/may/oecd-1980s-data-protection-principles.pdf (accessed 15 Dec 2017).

During the discussions in the OECD Expert Group the question was asked if the requirement of data quality relates to data protection.[27] External expert were divided on the correct classification[28] and it has been repeatedly established and pointed out that data quality was a general concept of computer science.[29] Nevertheless these Guidelines had a crucial role since they served as an inspiration for other legal frameworks such as the EU Data Protection Directive (Directive 95/46/EC).

CHAPTER 3. Data quality on European level

SECTION 1. – Art 5 of the Council of Europe Convention No. 108

On 28th January 1981 the Council of Europe adopted the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. Chapter II contains the basic principles of data protection. The data quality principle is established in Art. 5 (d) "that personal data undergoing automatic processing shall be [...] accurate and, where necessary, kept up to date". Art. 9 (2) provides specific exceptions when such derogations are allowed by the law of the party and is necessary for e.g. protecting state security or public safety. To ensure compliance with the principles established by national law, each country can establish appropriate sanctions and remedies. The Convention was the first intergovernmental data protection agreement to pave the way for a common European data protection regime. All signatory countries committed to enact national data protection regulations incorporating the principles laid down in the Convention. Although Germany was among the first countries to sign the Convention, it took until June 1985 to be ratified, so that the Convention entered into force on 1st of October 1985.

In 2001, the data protection convention was extended by an additional protocol obliging the signatory states to create independent supervisory authorities.

---

[27] It is explicitly laid down in the explanations of the guidelines, Explanatory Memorandum, p. 53.

[28] *Cf.* Gl. Gonzáles Fuster, *The Emergence of Personal Data Protection as a Fundamental Right of the EU*, Brussels, Springer Verlag, 1st ed., 2014, pp. 78 et seq.

[29] R. Clarke, "The OECD Data Protection Guidelines: A Template for Evaluating Information Privacy Law and Proposals for Information Privacy Law", http://www.rogerclarke.com/DV/PaperOECD.html (accessed 15 Dec 2017).

SECTION 2. – Art. 6 of the EU Data Protection Directive

In October 1995 the EU Data Protection Directive 95/46/EC was issued. It also contains, in Art. 6 (1) (d), the data quality principle. In contrast to the Convention, the Directive has a wider scope, as it is not limited to automatic processing of personal data. Art 6 (1) (d) determines that "Member States shall provide that personal data must be [...] accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that data which are inaccurate or incomplete, having regard to the purposes for which they were collected or for which they are further processed, are erased or rectified."[30]

Assessing the accuracy of data, time also plays a significant role. Data that was initially correct can become outdated if the information contained therein changes over time. It depends on the purpose, for which the date should be further processable if the data can be erased or requires rectification. This will always be the case if the processing result is factually incorrect without that particular information. Beside the problem of outdated information, incompleteness of data can also lead to it being incorrect, which is a quality problem. Art. 6 (2) establishes the obligation of the data processor to comply with the provisions of paragraph 1. Non-compliance with these provisions is not sanctioned, which seems to give the data principle the character of a recommendatory proposal.

Nevertheless, the European Court of Justice (ECJ) pointed out the high importance of the data quality principle, laid down in Article 6 of the Directive, in its Google decision.[31] The ECJ made very clear, that every processing of personal data must be measured by Article 6. "Even initially lawful processing of accurate data may, in the course of time, become incompatible with the Directive where those data are no longer necessary in the light of the purposes for which they were collected or processed", the court emphasized.[32]

Besides this big importance, it is surprising that the data quality principle has never been implemented completely and satisfactory into German

---

[30] *Official Journal* L 281, 23 Nov 1995 pp. 0031-0050.

[31] *Cf.* European Court of Justice, *Österreichischer Rundfunk et al. v. Rechungshof*, C-465/00, C-138/01 and C-139/01, EU:C:2003:294, Judgment of 20 Mar 2003, ref. 65; European Court of Justice, *ASNEF and FECEMD v. Administración del Estado*, C 468/10 and C 469/10; EU:C:2011:777, Judgment of 24 Nov 2011, ref. 26; and European Court of Justice, *Worten v. ACT*, C 342/12, EU:C:2013:355, Judgment of 30 May 2013, ref. 33.

[32] European Court of Justice, *Google v. Spain*, C-131/12, EU:C:2014:317, Judgment of 13 May 2014, ref. 93.

national law, i.e., the German Federal Data Protection Act (BDSG)[33]. However, one of the first countries within the EU to implement the EU Data Protection Principles was the United Kingdom in its Data Protection Act 1998. Prior to that, the Data Protection Act of 1984 had been in force. It was only applicable for automated processing of personal data.[34] The Data Protection Act 1984 has been replaced by the Data Protection Act 1998, which is based on the EU Data Protection Directive. The Data Protection Act 1998 is the legal framework of the British Data Protection Law and concrete legal requirements are established in statutory instruments and Codes of Practice.[35] The Data Protection Act 1998 determines in Schedule 1 (1) (4) eight data protection principles. The fourth principle stipulates that "personal data must be accurate and kept up to date"[36] which corresponds to the provisions of data quality established in Art. 6 (1) (d) of the EU Data Protection Directive. The definition of data accuracy is not contained in the Data Protection Act, but section 70 (2) describes personal data as inaccurate if it is "incorrect or misleading as to any matter of fact".[37] In the legal examination of a violation of the established data quality principle, the Data Protection Act distinguishes between data compiled by the data processor on his own or data which is provided by the data subject itself or is obtained by third parties. In the first case the data processor must take all reasonable steps to make sure that the information he gathered from his own sources is correct. If the use of the personal data could have serious implications he should show particular prudence. The Data Protection Act has a special regulation regarding inaccurate information provided by the data subject itself or received from third parties. In that case the data processor does not infringe the established data quality principle if the provided data have been recorded in an accurate manner, if he has taken adequate steps in regard to the purpose of the processing to assure the accuracy of the data in concern, and if the data make it evident, that the data subject has informed the data processor about the inaccuracy of the data.[38] The answer to the question of what reasonable steps are to be taken by the data processor to ensure

---

[33] Act amending the BDSG (Federal Data Protection Act) and other laws of 22 May 2001 (*Federal Law Gazette* I p. 904 et seq.).

[34] I. J. LLOYD, "Data Protection in the United Kingdom", in . KILIAN (ed.), *EC Data Protection Directive*, Darmstadt, Toeche-Mittler, 1st ed., 1997, p. 87.

[35] T. WESSING, "An overview of UK data protection law", https://united-kingdom.taylorwessing.com/uploads/tx_siruplawyermanagement/NB_000168_Overview_UK_data_protection_law_WEB.pdf (accessed 15 Dec 2017).

[36] In the Data Protection Act 1984 was established in Sch. 1 Pt. 1 para. 5, that "personal data shall be accurate and, where necessary, kept up to date".

[37] P. CAREY, *Data Protection in the UK*, London, Blackstone, 1st ed., 2000, p. 32.

[38] Schedule 1 (2) (7) Data Protection Act 1998.

that personal data is accurate depends, on the one hand, on the nature of the personal data concerned and, on the other hand, on the intention of use and the effects that inaccuracy of that information might have for the data subject.[39]

The fact that the Data Protection Act does not establish an absolute obligation to maintain the accuracy of personal data was accentuated in the case of *Smeaton v Equifax Plc* in 2013 by the UK Court of Appeal. Only reasonable steps must be taken to assure data quality. What is reasonable needs to be judged on a case by case basis. The Court also made it very clear that violation of the fourth data protection principle does not provide for a parallel duty in tort law.[40]

The EU Data Protection Directive was implemented into the Austrian Data Protection Act (*Datenschutzgesetz* 2000 – DSG 2000) in 2000.[41] The Austrian Data Protection Act (ADPA) implemented the data quality principle of Art. 6 (1) (d) of the European Data Protection Directive in its section 6 (1) (4). But Austria already had a Data Protection Law since 1978.[42] The data quality principle was incorporated to the ADPA as a consequence of the implementation of the EU Data Protection Directive. Section 6 (1) (4) ADPA determines that data may only be used in such a way that it is factually correct and, if necessary, updated with regard to the intended purpose. Section 27 (1) ADPA lays down the obligation of the data controller to rectify or delete inaccurate data of their own accord, as soon as he becomes aware of its inaccuracy or the inadmissibility of their processing. The same obligation of rectification or deletion faces the data controller upon a justified request by the affected data subject. Incompleteness of personal data may also lead to inaccurate information. But, as a matter of fact, a rectification is only necessary if the incurred data are important for the purpose of the data use. The incompleteness of the data used entitles rectification only if it leads to incorrect aggregated information. What is particularly interesting, is that under ADPA not only

---

[39] https://ico.org.uk/for-organisations/guide-to-data-protection/principle-4-accuracy/ (accessed 15 Dec 2017).

[40] Court of Appeal (United Kingdom), *Smeaton v Equifax Plc*, ECWA Civ 108, Judgment of 20 Feb 2013, http://www.bailii.org/ew/cases/EWCA/Civ/2013/108.html (accessed 15 Dec 2017).

[41] The Data Protection Act is a Federal Act which contains in its first section a "fundamental right to data protection" which has the statute of a constitutional law. In Austria, all individual rights which are contained in statutory regulations are strengthened by constitutional protections (L. PRAKKE, C. KORTMANN, "Constitutional Law of 15 Member States", 2004, p. 67).

[42] W. KOTSCHY, "The Implementation of the Data Protection-Directive 95/46/EG in Austria", in W. KILIAN (ed.), *EC Data Protection Directive*, Darmstadt, Toeche-Mittler, 1st ed., 1997, p. 15.

natural persons but also moral persons or associations are protected and therefore have the right under section 1 (3) (2) ADPA of rectification of inaccurate personal data in case of their processing.[43] Hence the scope of the data quality principle in Austria is broader than in the UK, because personal data of moral persons and associations fall within the application of this principle under the circumstances mentioned above.

SECTION 3. – Data quality in the GDPR

§ 1. Article 5 (1) lit. d GDPR as laid down principle of data quality

As in Article 6 (1) (d) of the EU Data Protection Directive the principle of data quality is also laid down as a principle in the GDPR. Article 5 (1) (d) GDPR uses the term of *data accuracy*. Significant in view of the principle are the purposes the data are stored for.[44] In some cases less accurate data may be sufficient, while in others high accuracy is needed to fulfill the purpose.

The wording in the GDPR leads to uncertainties. As mentioned above, the English version uses the term *accurate*, not only in Article 5 (1) (d) GDPR, but also in Recital 39 GDPR. For example in the German translation the term *richtig* (*correct*) is used and suggests the criterion applies to facts only and that facts may be classified in dual categorization as *accurate* (or *correct*) and *inaccurate* (or *incorrect*). The English term, however, is much more complex than its German translation and points out that data quality is not a question of *right* or *wrong*. Big Data is concerned with correlations and probabilities, and is thus not suited to dualistic assertions of truth. The term *accurate* comprises precision in the mathematical sense and purposefulness. The term is a central definition in modern ISO-standards,[45] underlining its roots in engineering sciences and early computer science. In this context, the German term for *correct* can be found in the above-mentioned special rules for statistics authorities and

[43] The EU Data Protection Directives scope was the protection of data regarding natural persons. The ADPA of 1978 protected also moral persons. But as explicitly postulated in recital 24, the protection of data on moral persons does not fall under the scope of the Directive. It was possible for Austria to uphold the protection of such data under its national law (KOTSCHY, *ibid.*, pp. 16 et seq.).

[44] T. HERBST in: J. KÜHLING, B. BUCHNER (eds), *Datenschutz-Grundverordnung*, München, C.H. Beck, ed., 2011, Article 5 ref. 62.

[45] ISO 5725-1:1994.

aviation organizations and thereby should be understood in the sense of *rather accurate*.

§ 2. The rights of data subjects with other interpretation of data quality

A central matter of the GDPR are the rights of data subjects, Articles 12-23 GDPR. In relation to the principle of data quality "the data subject shall have the right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her", Article 16 GDPR. In cases of controversy regarding to the accuracy of data the data subject may obtain the restriction of the processing "for a period enabling the controller to verify the accuracy of the personal data", Article 18 (1) (a) GDPR.

These wordings indicate an interpretation of data as *right* or *wrong*. Therefore, data accuracy has to be interpreted in two ways. On the one hand accuracy regarding Article 5 GDPR has to be interpreted in a technologically-relative way. On the other hand, the rights in Article 16 and 18 (1) (a) GDPR require that data must be treated as *correct* or *incorrect* in the ontological, bipolar sense. For example, address data of a data subject can be clearly wrong or not up-to-date anymore.

Data that cannot be classified as accurate or inaccurate (*non liquet*), is not inaccurate in the ontological bipolar sense and the rights in Article 16 and 18 (1) (a) GDPR are not applicable.

§ 3. Interpretation regarding the systematic nature of the GDPR

However, the relationship to Article 6 GDPR – determining the lawfulness of processing – is unclear. The requirement of data accuracy could be used as permission in terms of Article 6 (1) (f) GDPR. It could be interpreted as Article 5 GDPR requiring data to be up-to-date at any time as a legitimate interest in data processing. Other interpretations seem possible, too.[46]

The systematic point of view leads to the question of application of Article 5 GDPR in cases of non-compliance that do not have a negative impact on the data subject. In our opinion, the GDPR has to be

[46] J. Ph. Albrecht, Fl. Jotzo, *Das neue Datenschutzrecht der EU*, Baden-Baden, Nomos, 1st ed., 2017, p. 54.

interpreted restrictively[47] and Article 5 GDPR has to be reduced from a teleological point of view. As a result, data accuracy is only necessary if the non-compliance had a negative impact. The fact that the violation of Article 5 (1) (d) GDPR can be the basis of massive fines (Art. 83 (5) (a) GDPR) strengthens this restrictive interpretation. However, this does not mean that the abstract strict liability tort becomes a concrete one. This would be an interpretation against the wording of Article 5 (1) (d) GDPR.

§ 4. The importance of sanctions for violation of data quality

Article 83 (5) (a) GDPR marks a big reform in the data protection legislative, as the violation of principles on data regulation (Article 5 GDPR) – and therewith of the data quality principle – will entail severe fines. Yet the principles from the EU Data Protection Directive did not contain any sanctions.[48] Furthermore, Article 84 provides the possibility for Member States to regulate further criminal sanctions, whereby additional sanctions may be unattractive in the perspective of the Member States as they might lead to a disadvantage for the settlement of companies compared to other Member States.[49]

The new rules come with fines, which can be 4 % of annual sales for violating the provision on data quality. For that reason it is important, that the new regulations of the GDPR can be defined precisely. This gains special significance with regard to Article 103 (2) GG, which determines a high standard of precision before fines can be imposed.[50] Similar requirements are laid down in Article 49 (1) of the Charter of Fundamental Rights and Article 7 (1) of the European Convention on Human Rights. As mentioned above, especially the criterion of *factual accuracy* and its translations is vague. Therefore the fact that the GDPR criminalizes the violation of data quality is at the least a difficult issue with regard to the principle of precision.

---

[47] I. Anastasopoulou, *Deliktstypen zum Schutz kollektiver Rechtsgüter*, München, C.H. Beck, 2005, pp. 63 et seq.; E. Graul, *Abstrakte Gefährdungsdelikte und Präsumtionen im Strafrecht*, Berlin, Dunker und Humblot, 1991, pp. 144 et seq.; W. Gallas, "Abstrakte und konkrete Gefährdung", in: H.Lüttger (ed.) et al., *Festschrift für Ernst Heinitz zum 70. Geburtstag*, Berlin, De Gruyter, 1972, p. 171.

[48] See Art. 5 (1) (d) version from 11 Jun 2015, "Personal data must be accurate and, where necessary, kept up to date".

[49] P. Nemitz in: E. Ehmann, M. Selmayr (eds), *DS-GVO*, München, C.H. Beck, 2017, Art. 84 ref. 2.

[50] German Federal Constitutional Court, Case No 2 BvL 11/85, Vol. 50 1987 *NJW* 3175, Judgment of 6 May 1987, p. 341.

Additionally, there is a risk that the supervisory authority expands to a super-authority. The risk is caused by the complex term *accurate* in combination with the broad term of personal data as defined in Article 4 (1) GDPR. The data protection supervisory authority is in no position to judge the mathematical-statistical validity of scoring procedures. They have never been responsible for supervising such requirements. In that case the data protection supervisory authority would obviously have to employ mathematicians to check the validity, which would lead to additional administrative costs. That issue raises the question of whether data quality is (or should be) actually a field of data protection law.

CHAPTER 4. Relation of data quality to data protection?

As described above, the principle of data quality has already found its way into data protection law. However, it seems questionable whether data protection law is the right place to address the issue of data quality as it affects not only consumers but business people as well.

According to the data quality pioneers *Wang* and *Strong*, information quality can be defined as information that is fit for use.[51] But how do we determine which information is fit for use? And how does this intention – more similar to warranty law or civil law in general – fit to the idea of protecting customers and their informational self-determination? A good example for the importance of data quality is credit scoring. The mortgage industry is an information-intensive industry, which uses structured and unstructured data for the evaluation of borrowers.[52] How incorrect information regarding the solvency of a company can have a disastrous impact on companies became evident in the German case *Kirchgruppe v. Deutsche Bank*, for example.[53] This leads to the conclusion, that in credit

[51] R. T. Wigand, J. Wood, Y. Yiliyasi, "Information Quality Issues in the Mortgage Banking Industry", https://pdfs.semanticscholar.org/31d0/fc8af6bb4c6d59e8420cae7d807bc812e498.pdf (accessed 15 Dec 2017).

[52] R. T. Wigand, J. Wood, Y. Yiliyasi, "Information Quality Issues in the Mortgage Banking Industry", https://pdfs.semanticscholar.org/31d0/fc8af6bb4c6d59e8420cae7d807bc812e498.pdf (accessed 15 Dec 2017).

[53] For this purpose, German Federal Supreme Court, Case No. XI ZR 384/03, Vol. 12 2006 *NJW* 830, Judgment of 24 Jan 2006; P. Derleder, "Das Milliardengrab", *NJW*, 2013, p. 1786 et seq.; Cl. Höpfner, M. Seibl, "Bankvertragliche Loyalitätspflicht und Haftung für kreditschädigende Äußerungen nach dem Kirch-Urteil", *BB*, 2006, pp. 673 et seq.

scoring – which is regulated in German data protection law in section 28b BDSG (cf. section 31 of the new BDSG) – the essential question is not if the use of personal data was allowed but if the procedure was accurate and leads to an adequate result. Data protection on the other hand does not address those questions of data accuracy. Insofar section 28b BDSG is incorrectly located in the data protection legislation.

Moreover, consumer protection law does not find a satisfactory answer to questions of data quality. This is due to the fact that bad data quality affects not only consumers. It is incomprehensible why only natural persons should be affected by inaccurate data. Data quality is also crucial for the success or failure of companies. Especially in times of Big Data and data being valuable assets for many companies, the guarantee of data quality and protection against the misuse of Big Data tools becomes more important for moral persons, too. In order to react to that fact Austria has extended their data protection also to moral persons and therefore amplified the scope of protection also to enterprises and associations. Although this solution does not solve the problem that data quality should be an issue of civil law in general, it can be seen as an approach that takes into account the interest of companies in adequate data.

It can be seen quite clearly that data protection provisions do not apply in all data quality cases and that some data quality regulations (Art. 6 (1) (d) of the European Data Protection Directive, section 28b BDSG, Art. 5 (1) (d) GDPR) are located incorrectly in data protection legislation. Therefore data quality should be an issue of civil law in general to pave the way for companies in the age of Big Data and provide effective protection.

Conclusions

Through the provisions from the United States as well as developments from the European Data Protection Directive to the General Data Protection Regulation, the growing relevance of data quality becomes clear. However, adequate quality standards for data accuracy and veracity[54] can only be guaranteed by effective mechanisms. Thus the EU Directive on the European level was a cornerstone, although its provisions were not enough.

[54] See overview "Four V's of Big Data" (Volume, Variety, Velocity und Veracity) in S. MOHANTY, "The Four Essential V's for a Big Data Analytics Platform", *Dataconomy-Online*, http://dataconomy.com/the-four-essentials-vs-for-a-big-data-analytics-platform/ (accessed 15 Dec 2017).

Already in 2006, Yves Poullet criticized that the EU Data Protection Directive led to different interpretations and data protection standards within the Member States. From this point of view the GDPR can be seen as a progressive step towards a more harmonized data protection law. In addition to that, data quality as a basis for fines grows more effective within the GDPR. As it has been shown, new issues have arisen because there are no globally valid and recognized industry standards for data quality. We are still far from a harmonization and standardization regarding data quality standards. In this regard, the data protection supervisory authorities should take the new approach of criminal sanctioning of data quality very cautiously and carefully.

Another point can be seen as even more important for the data quality principle, as Yves Poullet mentioned in 2006: "The machine is the problem: the solution is in the machine"[55]. Already in 2006 when Big Data played a far less essential role than today, Yves Poullet emphasized the importance of problems concerning technical procedures. One could say his call was heeded with regard to the GDPR. The principles of *Privacy by Design* and *Privacy by Default* were laid down in Article 25 GDPR. As it is impossible to anticipate all technical developments in legislative processes, the machine could be used as a tool to solve its own problems on a legislative basis.

This approach demonstrates another aspect: The origin of the idea of data quality is the machine. Especially because of powerful machines and new tools for data analysis it is a concern to ensure that data meet expectations. On the one hand new questions of data protection were raised; on the other hand new business areas have a pressing need for data quality as basis for effective Big Data tools.

According to the visionary words of Yves Poullet, the importance of the machine should not be neglected. To meet these requirements the GDPR can be seen only as a first step in the right direction. Reflecting the importance of machines and Big Data processes, only general civil law legislation can be the answer.

[55] Y. POULLET, "EU data Protection policy: The Directive 9/46/EC: Ten years after", *Computer Law & Security Report*, 2006, p. 206.