CEJ
COMPLIANCE
ELLIANCE
JOURNAL

COMPLIANCE TECH

Tools for a modern compliance framework

Micha-Manuel Bues

AUTHOR

Dr. Micha-Manuel Bues, M.JUR. (Oxford) is a lawyer at Gleiss Lutz since 2013. He advises on German and European antitrust law as well as on all aspects of European law. He is, amongst others, specialized in antitrust compliance programs. He also advises and lectures on the use of legal technology and runs the blog www.legal-tech-blog.de. Micha-Manuel Bues studied at the universities of Passau, Bonn, Cologne and Oxford.

Data is everywhere. The volume, variety and velocity of data coming into companies have reached unprecedented levels. It is estimated that around 5 exabytes (= 5 billion gigabytes) of data are created each day, and this number is doubling roughly every 2 years.[1] Due to the sheer amount, companies do not know which potential compliance risks are "hidden" in the data. Companies are also confronted with more regulations, tightened enforcement and heightened global competition. This gives rise to new challenges for an effective compliance framework. However, surveys indicate that companies (depending on the industry and size) are relatively slow to adopt modern technologies to address these challenges. Law firms and lawyers are also reluctant when it comes to modern data analytics.

The only way to proactively or reactively investigate or prevent compliance issues in an era of data is to understand complex data sets from multiple sources within a company. In other industries, data analytics[2] is frequently used to uncover hidden patterns, unknown correlations and other useful information. Using data analytics, data scientists and others can analyze huge volumes of data that conventional analytics and business intelligence solutions can't touch. The emerging field of "Compliance Tech" tries to utilize data analytics in conjunction with subject matter professionals to bring compliance detection and prevention into the 21th century. Compliance Tech is used to proactively seek opportunities to detect and prevent fraud, waste and abuse to ensure compliant behavior. Compliance Tech aims to make the invisible visible. Compliance Tech provides the toolset to spot patterns or trends in data that are invisible to the naked eye. This is achieved by using different techniques (which will be explained later in greater detail) which helps to retrieve, organize, structure and streamline data to make invisible patterns become visible.

Compliance Tech brings compliance investigation or prevention to a new level. It allows for anomaly detection, clustering and risk ranking through a statistical-based analysis. This approach ensures better results compared to a "traditional" rule-based approach (matching, grouping, ordering, jointing and filtering etc.) when deployed over large data sets. Data visualization and test mining are, for example, superior to traditional keyword searching. Without these tools, companies remain dependent on human identification of risks and violations, whether flagged by employees, hotline tips, whistle-blowers or government auditors. Their compliance efforts often consist only of training employees to spot misconduct, and in setting aside financial reserves to fund expensive, after-the-fact investigations by outside counsel.

This article seeks to present the technologies, methodologies, processes and practices

1 Cf. VINCENZO MORABITO, BIG DATA AND ANALYTICS 105 (2015).

2 THOMAS A. RUNKLER, DATA ANALYTICS: MODELS AND ALGORITHMS FOR INTELLIGENT DATA ANALYSIS (2015), provides a good introduction into this field.

behind Compliance Tech and will give practitioners, especially lawyers, a short overview of this rapidly emerging but still relatively unknown field. The article does not try to give a comprehensive overview of the subject but intends to familiarize compliance practitioners with some of the fundamentals of data analytics. Compliance Tech forms part of a compliance program within a company. A compliance program will be defined here as an internal program and decision policy made by a company in order to meet the standards set by government laws and regulations.[3]

Most of the technologies and methodologies deployed in Compliance Tech require **big data**[4]; a buzzword that describes a volume data so large that it is difficult to process using traditional database and software techniques.[5] An important distinction to bear in mind is between structured and unstructured data.[6] Structured data which is stored in precisely defined and described data fields. A typical example is a customer database in which each record consists of a name, address, birth date, etc. Structured data have a clear model and description and are therefore easily stored, processed and analyzed. Conversely, unstructured data do not have a precisely defined structure. This category may include images, videos, websites or content of e-mail and/or other communications. Unstructured data constitutes the absolute majority of generated data.

There are many techniques that draw on disciplines such as statistics and computer science (particularly machine learning) that can be used to analyze the structured and unstructured datasets. The following non-exhaustive list entails the most commonly used techniques in Compliance Tech. Not all of these techniques strictly require the use of big data; some can be applied to effectively smaller datasets. The techniques and methodologies are presented here in an order in which they would be typically used. Depending on the actual compliance framework required, different techniques may be used or combined to cater for the particular needs of a company.

A common first step in deploying Compliance Tech is **data acquisition**.[7] This aims to ensure a stable transfer of internal and external data and convert the data into a format suitable for further analysis. **Data mapping** serves as the initial step in data inte-

---

3 Cf. JULIA STEHMANN, COMPLIANCE-MANAGEMENT 6 ff. (2011).

4 VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW LIVE, WORK, AND THINK (2013) gives a non-scientific overview of the opportunities and challenges of big data.

5 Cf. JAMES R. KALYVAS & MICHAEL R. OVERLY, BIG DATA: A BUSINESS AND LEGAL GUIDE 1 (2015); ANIL AGGARWAL, MANAGING BIG DATA INTEGRATION IN THE PUBLIC SECTOR 73 (2015).

6 Cf. MICHAEL BRACKETT, DATA RESOURCE DATA: A COMPREHENSIVE DATA RESOURCE UNDERSTANDING 14 (2014).

7 MAURIZIO DI PAOLO EMILLIO, DATA ACQUISITIONS SYSTEMS (2013) gives a detailed introduction into this area.

gration.[8] Data mapping is the process by which different data models are linked. This uses a defined set of methods to characterize the data in a specific definition. The technique involves evaluating data values in different data sources, as well as automatically and simultaneously discovering complex mappings between the sets. Data mapping is also used to consolidate multiple databases into a single database.

Link analysis is a technique used to evaluate the relationships or connections between various types of objects (nodes), including people, organizations and transactions.[9] Link analysis is a kind of knowledge discovery that can be used to visualize data, allowing for better analysis, especially in the context of links (web links or relationship links). Link analysis might be able to detect or establish "hidden" relationships within a data set.

Social network analysis (SNA) is the process of quantitative and qualitative analysis of a given social network. SNA measures and maps the flow of relationships (ties) and relationship changes between entities.[10] Simple and complex entities (nodes) include websites, computers, animals, humans, groups, organizations and nations. Along with link analysis, SNA can, for example, help to identify related parties, conflict of interests, corruption and bid rigging.

Text mining is the analysis of data contained in natural language text.[11] The application of text mining techniques to solve business problems is called text analytics. Text analytics software can help by transposing words and phrases in unstructured data into numerical values which can then be linked with structured data in a database and analyzed with traditional data mining techniques.

Data mining is the process sorting through data to identify patterns and establish relationships.[12] Data mining parameters include:

- Association, i.e. identifying associations between events.
- Sequence or path analysis, i.e. identifying patterns where one event leads to another later event.
- Classification is a set of techniques used to identify the categories in which new data points belong, based on a training set containing data points that have al-

8 Cf. QAMAR SHAHBAZ, DATA MAPPING FOR DATA WAREHOUSE DESIGN Chapter 3 (2015).

9 Cf. B. KIRWAN & L. K. AINSWORTH, A GUIDE TO TASK ANALYSIS 116 ff. (1992).

10 Cf. IAN MCCULLOH & HELEN ARMSTRONG, SOCIAL NETWORK ANALYSIS WITH APPLICATIONS Introduction (2013).

11 Cf. STÉPHANE TUFFÉRY, DATA MINING AND STATISTICS FOR DECISION MAKING 627 (2011).

12 Cf. STEPHAN KUDYBA & RICHARD HOPTROFF, DATA MINING AND BUSINESS INTELLIGENCE: A GUIDE TO PRODUCTIVITY 37 (2001).

ready been categorized.[13]

- Clustering is used to place data elements into related groups without advance knowledge of the group definitions.[14]

Discovering patterns in data that can lead to reasonable predictions of future trends is known as **predictive analytics**.[15] The central element of predictive analytics is the predictor, a variable that can be measured for an entity to predict future behavior. Multiple predictors are combined into a predictive model, which can be used to forecast future probabilities with an acceptable level of reliability.

Data visualization is a general term for tools which help to understand the significance of a particular data set by placing it in a visual context.[16] As a result of the visualization, patterns, trends and correlations can be exposed and recognized more easily. State-of-the-art data visualization software goes beyond the standard charts and graphs used in Excel. Data visualization entails more sophisticated tools such as infographics, dials and gauges, geographic maps, sparklines, heat maps, and detailed bar, pie and fever charts. One of the most widely used visual techniques is a tag cloud. A tag cloud is a stylized way of visually representing rates of occurrences of words used to described tags. The most popular topics are normally highlighted in a larger, bolder font. Data visualization helps us to absorb large pieces of information more efficiently.

Combining the described methods significantly enhances the overall probability of detecting and preventing compliance incidents within a company. Law practitioners should a have a good understanding of these methods to ensure the best results for their clients. However in practice, there might be several challenges to overcome if a company, compliance department or law firm decides to use Compliance Tech.

One challenge with Compliance Tech is that companies or compliance teams need to understand how the tools work in practice. The challenge is to find the team with the right skillset. An ideal Compliance Tech team should be a combination of compliance experts with a legal background, and data scientists. All team members should be capable of working both with new technologies, and interpreting data to find meaningful compliance insights. However, in reality, members of compliance teams normally only

---

13 Cf. THOMAS A. RUNKLER, DATA ANALYTICS: MODELS AND ALGORITHMS FOR INTELLIGENT DATA ANALYSIS 85 (2015).

14 Cf. THOMAS A. RUNKLER, DATA ANALYTICS: MODELS AND ALGORITHMS FOR INTELLIGENT DATA ANALYSIS 103 ff. (2015).

15 COLLEEN MCCUE, DATA MINING AND PREDICTIVE ANALYSIS (2014); DEAN ABBOTT, APPLIED PREDICTIVE ANALYTICS (2014), give a good overview over the techniques used in predictive analytics.

16 Cf. Evan F. Sinar, *Chapter 5 (Data Visualization)*, in Big Data at Work 115 (Scott Tonidandel et al eds., 2015).

have a legal or non-tech background. Without proper training, they tend to have problems understanding the underlying concepts and methodologies of Compliance Tech. It is then difficult to effectively supervise the data analytic process and to apply the right technology architecture and capabilities. This expertise could either be built in-house, or outsourced to a third party IT partner. As the technology landscape in the data world is evolving extremely fast it could be helpful to work with a strong and innovative technology partner who can help create the right IT architecture to efficiently adapt to changes in the landscape.

In addition, any introduction of Compliance Tech should be accompanied with a change management approach that includes an extensive communication effort. Many companies fail to recognize that new analytics often requires new behaviors.[17] For this reason, communication plays an essential role to educate, inform and explain a Compliance Tech approach within a company or legal department. Personal experience has shown that it takes a considerable amount of time to introduce the concepts of Compliance Tech and their benefits to employees, business stakeholders, management and IT teams. The management must be willing to change in order for the data and models to yield better compliance decisions.[18]

In addition, it is important to find the right applications for Compliance Tech. Sometimes companies are lured into thinking that running analytics on a very large set of data is data analytics. But data analytics only shows it best results when it's used on a concrete and meaningful compliance case. Therefore, it is essential that the compliance team identifies the right data and has a good understanding of the data structure within a company. The sheer volume of information, particularly from new sources such as social media, is growing rapidly. Bigger and better data give companies a more panoramic and granular view of their business environment and potential compliance pitfalls. This all makes it more difficult to detect the right data or to see the potential value of data. Often the existing IT architecture may prevent the integration of stored information, and managing unstructured data often remains beyond traditional IT capabilities. It is therefore important to ensure an adequate IT infrastructure when using Compliance Tech.

Leveraging big data often means working across multiple disciplines such as IT, engineering, finance and procurement, and the ownership of data is fragmented across these disciplines. Addressing these organizational challenges means finding new ways of collaborating across functions and businesses. In this regard, it might be also sensible to

---

17 Michael Schrage, *Why your analytics are failing you*, Harvard Business Review (Apr. 8, 2014, 11:49 AM), https://hbr.org/2014/04/why-your-analytics-are-failing-you.

18 Dominic Barton & David Court, *Making advanced analytics work for you*, Harvard Business Review (Oct., 2012, 11:53 AM) https://hbr.org/2012/10/making-advanced-analytics-work-for-you.

drive an integrated approach to data sourcing, model building, and organizational transformation.

Data privacy and data security laws are one area of law that any business using big data will have to take very seriously.[19] In addition to local, state, national, and, even international laws, there are many other potentially applicable standards and guidance's. If a company uses Compliance Tech, it needs to ensure that its use is consistent with the above rules and regulations. In addition, as with many technological endeavors, big data analytics is prone to data breaches. Any data provided to a third party IT partner could get leaked, and needs to be protected accordingly.

It would be naive to see Compliance Tech as a panacea to cure all the woes of a compliance program. Although Compliance Tech is a remarkable tool that can help to enhance compliance efforts within a company, it is important to bear in mind that (a) the setup of a technology enhanced compliance program will take a considerable amount of time and effort and (b) that Compliance Tech has its inherent limitations. However, if these are understood properly Compliance Tech tools will be invaluable in compliance programs of the 21th century.

[19] For more information refer to JAMES R. KALYVAS & MICHAEL R. OVERLY, BIG DATA: A BUSINESS AND LEGAL GUIDE 33 ff. (2015).