12.01.23

APPLICATION OF BIG DATA TECHNOLOGIES IN THE MEDIA & ENTERTAINMENT INDUSTRY.

BIG DATA THEORY & PRACTICE
GROUP COURSEWORK (ASSESSMENT 002)
TECHNICAL REPORT - PART A

GROUP 4:

FRANCES WIREKO - W1782829
BETTY MBOCHE - W1791383
UNYIME AKPABIO - W1845262
YOGITA GANAGE - W1929606
FEJIRO ENI - W1880623

WORD COUNT: 2781

Table of Contents

LIST OF FIGURES

INTRODUCTION

Companies are rapidly improving Big Data strategies to effectively use massive data collections, particularly in the media and entertainment industries. Big data is a collection of structured and unstructured data that organizations collect and store in order to extract useful knowledge for predictive modelling, pattern recognition, machine learning, and big data analytics projects.

People today are very excited about new content and media platforms where they can view massive amounts of data. Big Data provides opportunities for growth and long-term development in the media and entertainment industries. The media and entertainment industry combines and collects similar types of data from various sources in order to understand the viewer's needs and behaviour and upgrade in such a way that they become more famous and the viewers' favourite among all competitors.

Big data provides such giants with massive amounts of categorised data. Using basic information such as search history, ratings, social media followed, trends, bookmarks, gender, age, and so on, the organization can predict customers' interests and personalize their experience. Because industries can now retrieve data based on various criteria such as age, location, language, and so on, they can channel and improve customer behaviour in order to keep them engaged and attract loyal customers.

WHY MEDIA & ENTERTAINMENT

In many ways, the media and entertainment industry were among the first to adopt big data technology, allowing them to improve digital transformation and capitalize on data for an industry that is already available, as well as new data sources both inside and outside the organization.

Broadcasters, news organizations, publishers, cable companies, and gaming companies in the media and entertainment industry are adapting new business models for how they create new content, market, and distribute it. It's popular because customers can search for and access content from any location, at any time, and using any device. This increases the pressure to develop new digital production and multi-channel advertising and distribution strategies based on a thorough understanding of consumer preferences, behaviours, and usage. On the other hand, because consumers can easily switch from analogue to digital media or to a new platform, there is a need to keep them engaged by monetizing all content and identifying new products and services of interest to them.

Every day, thousands of new customers connect to this media and enterprises via music, media, movie, and television show apps, and it is critical for these entertainment apps to maintain their popularity and demand, as well as client soothsaying, to know what they like and recommend their favourite content.

HOW BIG DATA IS USED IN MEDIA AND ENTERTAINMENT

- Big data aids in engaging target customers by providing recommendations based on previous viewings. Different users receive different recommendations based on their interests and history on YouTube and OTT platforms such as Netflix and Amazon Prime, generating more revenue.

- Television broadcasts advertisements and shows accordingly, allowing them to earn more money from advertisements and provide more engaging experiences, such as advertisements for cookies, chocolates, and children's products that run in the afternoon when children return home from school.
- By offering bundled content services, the media and entertainment industry is rapidly changing and adapting different roles and modifying applications. Allowing customers to stay on one platform by offering multiple services on the same platform. For example, Disney launched Disney+ and granted permission for other platforms to host its films and shows with just one click.

PROBLEMS FACED IN THE INDUSTRY

The following are some of the limitations of big data in the media and entertainment industries:

- Consumer awareness and concern about how personal data is used have grown. There is regulatory uncertainty for European businesses that handle personal data, which could put them at a competitive disadvantage compared to, say, US companies that operate in a much more relaxed legal environment (Cavanillas, Curry and Wahlster, 2016).
- More data professionals are needed in Europe's labour market to manipulate big data applications, such as data journalism and product management (Cavanillas, Curry and Wahlster, 2016).
- Fear of piracy, as well as consumer disregard for copyright, may deter creative individuals and businesses from taking risks in launching new media and cultural products and services (Cavanillas, Curry and Wahlster, 2016).
- The content and data industries are dominated by large US players. Apple, Amazon, and Google dominate a wide range of industries, including music, advertising, publishing, and consumer media electronics (Cavanillas, Curry and Wahlster, 2016).

TYPES OF DATA USED

Thanks to social media and entertainment services, entertainment is now more distinct than it has ever been in the Digital World. With superfast internet and new technology, there is a greater demand for online entertainment. According to statista.com, the global video streaming market has experienced unprecedented growth, with video on demand revenue exceeding 86 billion US dollars in 2021, 1.2 billion Subscription Video On Demand (SVOD) subscriptions worldwide, and a record 44 percent of all TV viewing time spent watching content on streaming services as of the third quarter of 2021 (Most popular streaming brands 2022, no date).

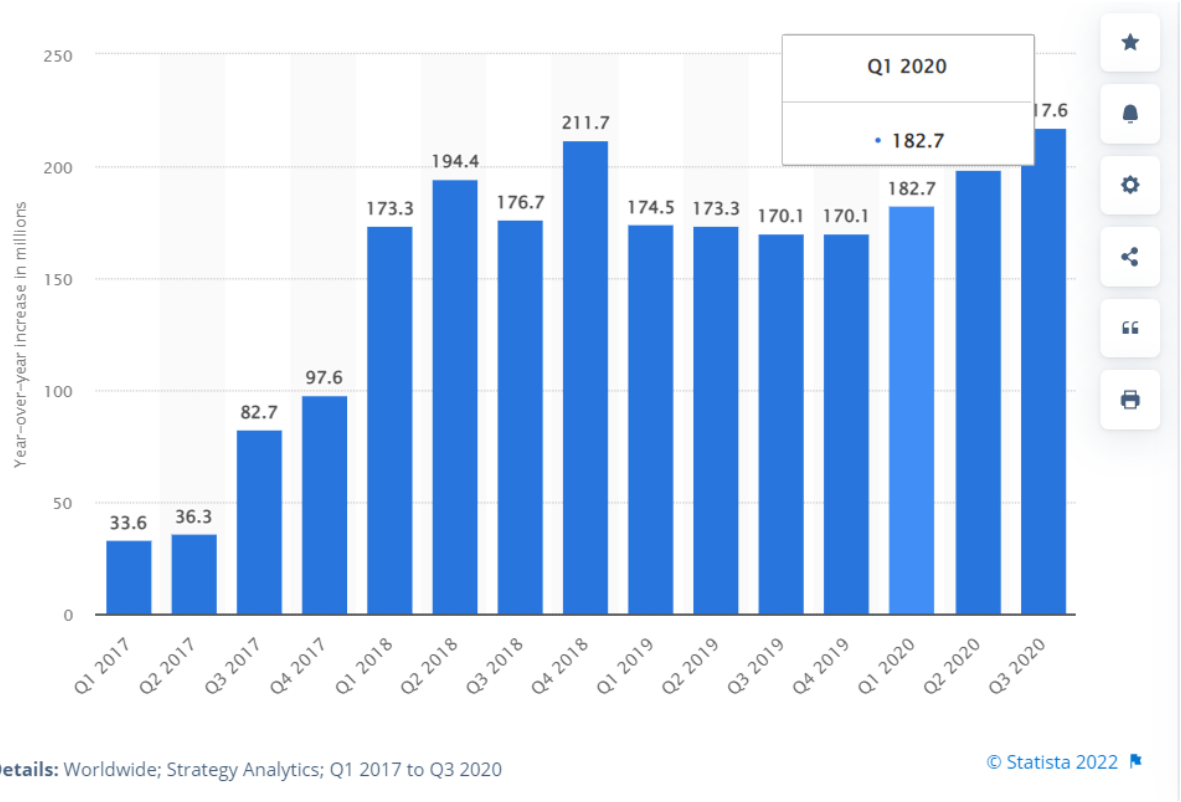

Figure 1: Number of subscription video-on-demand (SVOD) subscriptions worldwide

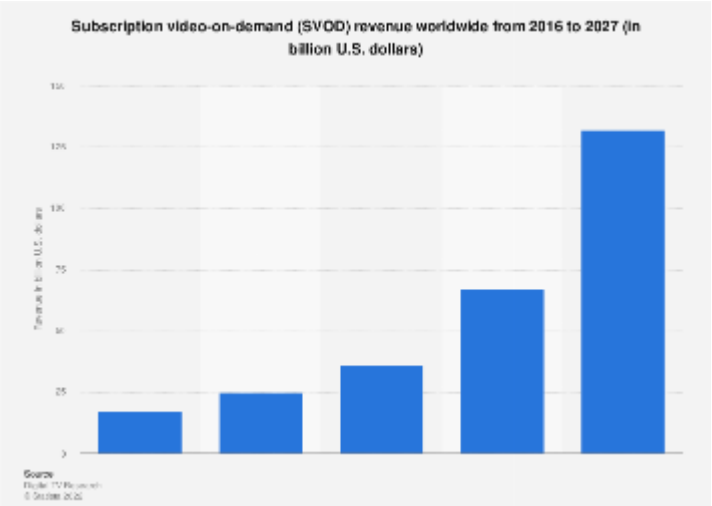

Figure 2: Subscription video-on-demand (SVOD) revenue worldwide from 2016 to 2027

According to data on worldwide digital video viewership released in 2021, there were over three billion internet users of all ages in 2020 who watched streamed or downloaded videos on any device at least once a month. This figure is expected to nearly triple to 3.5 billion by 2023 (Cavanillas, Curry and Wahlster, 2016).

The reality is that the more we stream, the more data about our habits, preferences, whereabouts, and so on is gathered online. Big data collection and analysis are changing as a result of the global digital transformation (How data and streaming are transforming entertainment as we know it, 2017). The amount of data generated by Internet traffic alone can be massive and used to generate useful

information. According to experts, video streaming would have accounted for 80% of all internet traffic by 2019 (Cavanillas, Curry and Wahlster, 2016).

The fundamentals of data types are the same across all big data processing architectures: data is first gathered from disparate information sources with the goal of being stored in scalable, big data-capable data storage (Event streaming technologies a remedy for big data's onslaught | TechTarget, no date). Three key elements are required to accomplish this:

- Protocols for gathering information from distributed data sources of any kind
- Data collection methods and tools for dispersed sources using various protocols
- Technologies that allow the frameworks' retrieved data to be saved indefinitely

CASE STUDY

SPOTIFY:

The popular music streaming service Spotify uses a big data analytics platform called Hadoop to gather enormous volumes of data from its more than 60 million active users globally. Spotify then uses the analysed information from user profiles, playlists, and other historical data to provide specific users with informed, tailored, and pertinent recommendations for further music they might like to listen to.

Big data and predictive analytics are also used by Spotify to improve the user experience. One such is the Grammy Awards, where each year, Spotify makes award predictions based on streaming data. Based on the outcomes of the 2013, 2014, and 2015 Grammy Awards, Spotify's predictions were largely accurate (How Media Companies are Using Big Data, 2015).

AMAZON PRIME:

Customers come first with Amazon Prime, which offers "exclusive access to music, movies, TV shows, and Kindle books." With no need for drones, Amazon's "Prime Music" and "Prime Instant Video" streaming services use big data to offer customers a wide range of personalised entertainment options on a variety of devices, instantly, anywhere, and at any time. They do this by using the same sophisticated analytics tools that made Amazon.com the one-stop-shopping behemoth that it is today (How Media Companies are Using Big Data, 2015).

TECHNOLOGIES USED IN THE INDUSTRY

Netflix uses data processing software and traditional business intelligence tools such as Hadoop and Teradata, as well as its own open-source solutions such as Lipstick and Genie, to collect, store, and process massive amounts of data. These platforms influence its decisions regarding what content to create and promote to viewers (How Netflix uses big data to create content and enhance user experience, 2019). It does not use a traditional data centre-based Hadoop data warehouse. It uses Amazon's S3 to warehouse its data, allowing it to spin up multiple Hadoop clusters for different workloads accessing the same data. It uses Hive for ad hoc queries and analytics in the Hadoop

ecosystem, and Pig for ETL (extract, transform, load) and algorithms (How Netflix uses big data to create content and enhance user experience, 2019).

Netflix then developed its own Genie project to assist in handling ever-increasingly huge data volumes as it scales. All of this points to one thing: Netflix is highly concerned with having a large amount of data and being able to process this data in order to understand exactly what its viewers desire (How Netflix uses big data to create content and enhance user experience, 2019).

The Netflix Genie Project

Genie manages jobs and resources for the Hadoop ecosystem in the cloud. Genie abstracts away the physical details of various (potentially transient) Hadoop resources in the cloud and provides a REST-ful Execution Service to submit and monitor Hadoop, Hive, and Pig jobs without requiring any Hadoop clients to be installed. Genie also provides a set of Configuration Services, which serve as a registry for clusters and their associated Hive and Pig configurations from the perspective of a Hadoop administrator (Blog, 2017).

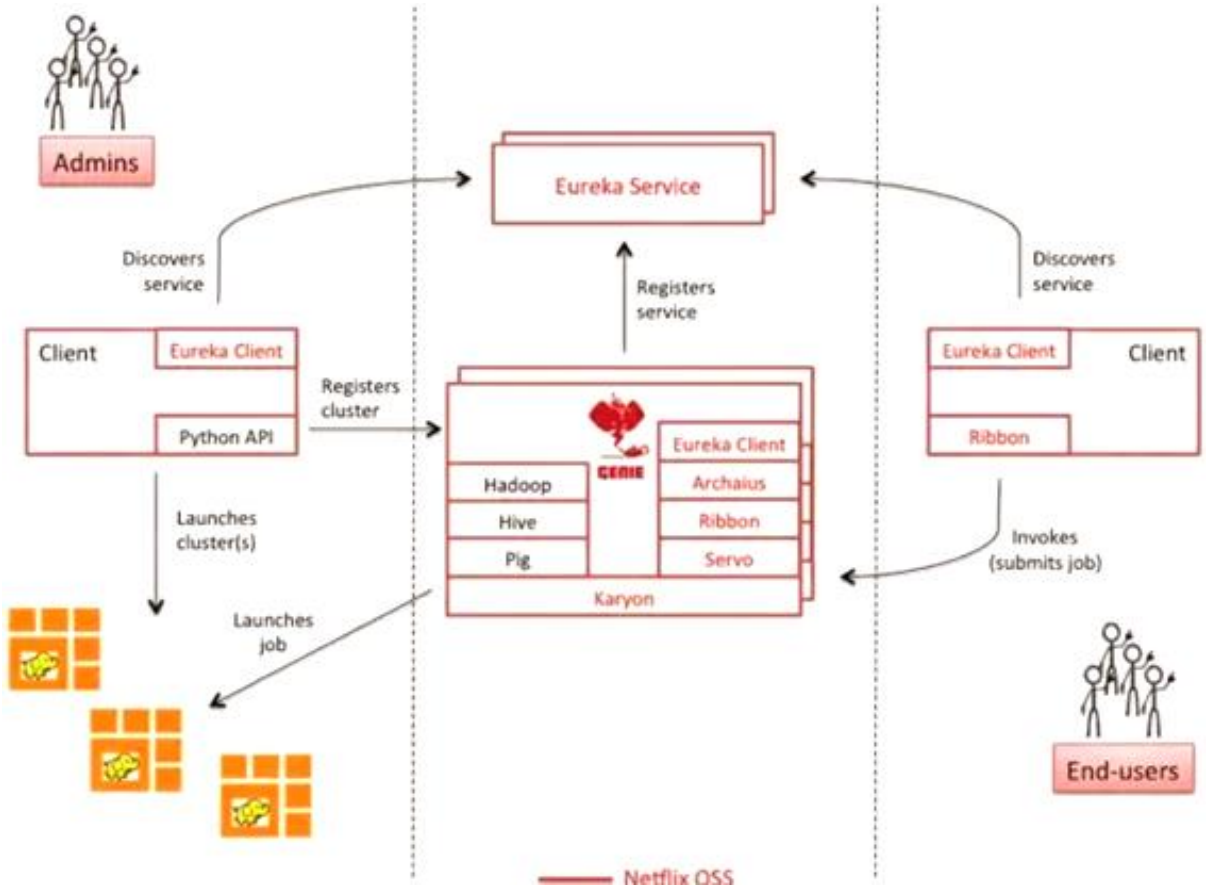

Figure 3: The architecture of genie

Genie itself is built on top of the following Netflix OSS components:

- Karyon, which provides bootstrapping, runtime insights, diagnostics, and various cloud-ready

hooks,

- Eureka, which provides service registration and discovery,
- Archaius, for dynamic property management in the cloud,
- Ribbon, which provides Eureka integration, and client-side load-balancing for REST-ful inter-process communication, and
- Servo, which enables exporting metrics, registering them with JMX (Java Management Extensions), and publishing them to external monitoring systems such as Amazon's CloudWatch (Blog, 2017).

RISKS FACED IN THE INDUSTRY

There are several risks that companies in the media and entertainment industry may face when using big data technologies:

- Data privacy and security: Large amounts of sensitive personal data, such as customer demographics and viewing habits, are frequently collected and processed by companies in the media and entertainment industries. If this data is not properly secured, unauthorized parties may gain access to it and use it for malicious purposes.
- Legal and regulatory compliance: The collection and use of personal data in the media and entertainment industry is governed by a number of laws and regulations, including the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act. Noncompliance with these laws can result in fines and reputational harm.
- Quality and accuracy of data: Inaccurate or incomplete data can lead to incorrect decisions and insights. Companies in the media and entertainment industries must ensure that the data they collect, and use is of high quality and accurately represents the trends and patterns being studied.
- Ethical considerations: The use of big data technologies in the media and entertainment industries raises ethical concerns, such as the possibility of biased algorithms and public opinion manipulation. Companies must be aware of these issues and take precautions to ensure that their use of big data is transparent and ethical.

SUCCESS & FAILURES OF TECHNOLOGIES UTILISED

Netflix is a well-known example of a company that has used big data technologies to successfully improve its operations. Among the specific examples of success are:

1. Improved recommendation algorithm: Netflix's "Cinematch" recommendation algorithm uses data on customer viewing habits and ratings to recommend content to individual users. The algorithm has proven to be extremely successful, contributing to Netflix's growth and customer retention.

2. Improved content creation and distribution: Netflix bases its content production and distribution decisions on data from viewer engagement and ratings. This has enabled the company to create popular shows such as "Stranger Things" and "Orange is the New Black," as well as reach new audiences worldwide.
3. Enhanced efficiency: Netflix optimizes its operations and makes data-driven decisions using data analytics. This has assisted the company in lowering its costs and improving its bottom line.

However, Netflix has encountered difficulties and failures as a result of its use of big data technologies. One example is the company's use of data on viewer habits to inform decisions about whether to renew or cancel TV shows. Some critics claim that this has resulted in the cancellation of shows with a smaller but dedicated audience, and that the data-driven approach does not always align with creative considerations.

DATA GOVERNANCE AND GDPR

A brand-new category of immersive entertainment known as "interactive entertainment" is being pioneered by several video-on-demand (VOD) companies. In reality, viewers' decisions at specific points in the plot of what they are watching affect it and result in a variety of personalised endings. The key legal issues are as follows:

- **Privacy:** While it may appear that interactive entertainment can amount to profiling, it is unclear whether data protection regulations cover viewers' interactions that go that far in entertainment. Because smart devices may collect massive amounts of metadata, electronic communications laws may apply to this processing. Additionally, viewers must be informed of the possibility of tracking and targeted profiling via detailed information notices and, if necessary, must consent to such processing.
- **Consumer law:** When it comes to tracking viewers, consumer law is concerned with initiatives by broadcasters and VOD platforms to take advantage of cross-device and online-to-offline surveillance of viewers' preferences and choices. This is because these behaviours may result in a reduction in the variety of information available to customers without any prior representation.
- **Data sharing:** Any method of exchanging data must be thoroughly investigated. Third parties providing audience measurement services, for example, must ensure that contracts adequately guarantee the privacy of the data they collect. Prior to collecting and transferring a large amount of personal data from viewers, it is also recommended to conduct a data protection impact assessment (in accordance with Article 35 GDPR).
- **Media (regulatory):** This is heavily influenced by how Member States implement the new Audio-visual Media Services Directive text. Furthermore, as a general observation, the EU Regulation No. 2017/1128 governing the cross-border portability of online content services in the EU market was eventually passed on May 24, 2017. This Regulation requires providers of pay-for-access online content services to provide subscribers with the same access to and

use of their services as those permanently residing in a Member State, including access to the same content on the same platform and number of devices, for the same number of users, and with the same range of functionalities.

- **Copyright protection** is not a new issue. For example, the gaming industry has long advocated that copyright rules be expanded globally to include the entirety of the computer code that creates the game and its features, just as they do for the entirety of an artwork or a work of literature. As "gamification" of entertainment spreads to traditional TV shows and VOD platforms, the question of whether "interactivity" may be protected in addition to the "conventional" copyright on the audio-visual work is raised" (EU DSM Copyright Directive: the use of protected content by online content-sharing service providers, no date).

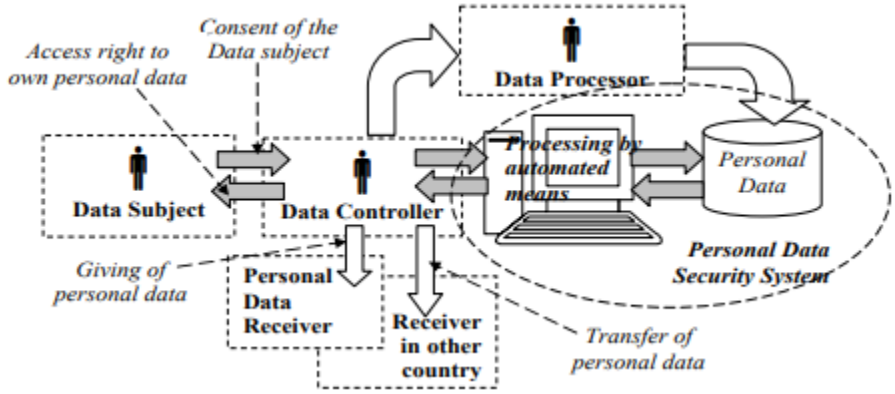

Figure 4: Organization of personal data processing

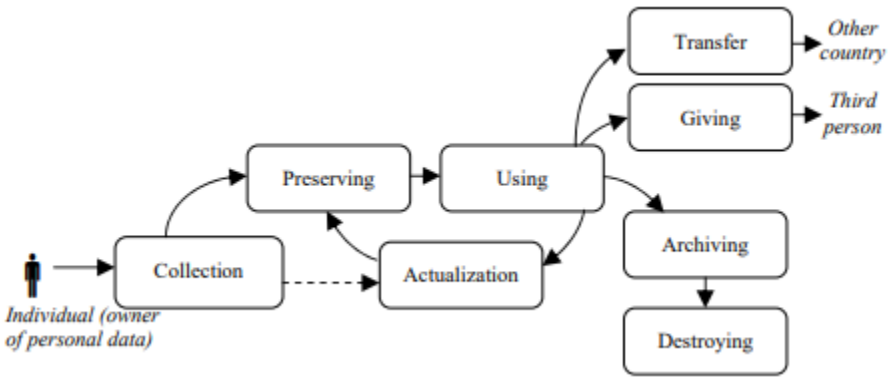

Figure 5: Life cycle of personal data processing

Spotify

The Spotify Privacy Policy explains how Spotify AB handles personal data. It is referred to as the "Policy" as of 12/12/2022.

It applies to the use of:

1. All Spotify streaming services as a user. i.e.
    - The usage of Spotify on any device
    - The customization of your user experience
    - The infrastructure needed to provide Spotify services
    - The integration of your Spotify account with other applications
    - Both our free or paid streaming options (each a 'Service Option')
2. other Spotify services which include a link to this Policy. These include Spotify websites, Customer Service and the Community Site

We'll refer to these as a whole moving forward as the "Spotify Service."
Occasionally, new services are created or provided. Unless otherwise specified, they'll also be governed by this Policy when introduced.

CONCLUSION

In conclusion, the use of big data technologies has transformed the media and entertainment industries. Companies can gain insights into consumer behaviour and preferences, personalise content and advertising, and optimise operations by allowing them to analyse massive amounts of data. The use of big data has also enabled the development of new business models, such as subscription and streaming services. While there are still obstacles to overcome, such as data privacy and ethical concerns, the future of big data technologies in the media and entertainment industry looks promising. Furthermore, the use of big data technologies has facilitated the growth of targeted advertising and personalised recommendations, which has resulted in increased revenue for media and entertainment companies. It has also enabled the development of more immersive and interactive consumer experiences, such as virtual and augmented reality content. We can expect to see even more innovative and impactful applications in the media and entertainment industry as big data technologies advance and become more widely adopted.

REFERENCES

Blog, N.T. (2017). Genie is out of the bottle! *Medium*. Available from https://netflixtechblog.com/genie-is-out-of-the-bottle-66b01784752a [Accessed 10 January 2023].

Cavanillas, J.M., Curry, E. and Wahlster, W. (eds.). (2016). *New Horizons for a Data-Driven Economy*. Cham: Springer International Publishing. Available from https://doi.org/10.1007/978-3-319-21569-3 [Accessed 10 January 2023].

Event streaming technologies a remedy for big data's onslaught | TechTarget. (no date). *Business Analytics*. Available from https://www.techtarget.com/searchbusinessanalytics/feature/Event-streaming-technologies-a-remedy-for-big-datas-onslaught [Accessed 10 January 2023].

EU DSM Copyright Directive: the use of protected content by online content-sharing service providers. (no date). Available from https://www.dentons.com/en/insights/articles/2019/april/12/eu-dsm-copyright-directive-the-use-of-protected-content-by-online-content-sharing-service-providers [Accessed 10 January 2023].

European Parliament approves the new Directive on Copyright. (no date). Available from https://www.dentons.com/en/insights/articles/2019/march/27/european-parliament-approves-the-new-directive-on-copyright [Accessed 10 January 2023].

How data and streaming are transforming entertainment as we know it. (2017). *Crayon Data*. Available from https://www.crayondata.com/how-data-and-streaming-are-transforming-entertainment-as-we-know-it/ [Accessed 10 January 2023].

How Media Companies are Using Big Data. (2015). *Qubole*. Available from https://www.qubole.com/blog/media-companies-using-big-data [Accessed 10 January 2023].

How Netflix uses big data to create content and enhance user experience. (2019). *ClickZ*. Available from https://www.clickz.com/how-netflix-uses-big-data-content/228201/ [Accessed 10 January 2023].

Most popular streaming brands 2022. (no date). *Statista*. Available from https://www.statista.com/statistics/1322864/streaming-brands-online-mentions/ [Accessed 10 January 2023].

Smart entertainment, privacy and the media industry. (no date). Available from https://www.dentons.com/en/insights/articles/2019/april/12/smart-entertainment-privacy-and-the-media-industry [Accessed 10 January 2023].

Romansky, R.P. (2014). Social Media and Personal Data Protection. *International Journal on Information Technologies & Security*, (4).