https://doi.org/10.53656/math2022-4-4-per

Educational Technologies
Образователни технологии

PERSONAL DATA PROCESSING IN A DIGITAL EDUCATIONAL ENVIRONMENT

Dr. Evgeniya Nikolova, Assoc. Prof.,
Dr. Mariya Monova-Zheleva, Assoc. Prof.,
Dr. Yanislav Zhelev, Assoc. Prof.
Institute of Mathematics and Informatics – Bulgarian Academy of Sciences (Bulgaria)

Abstract. New technologies provide innovative spaces for cooperation and communication between employers and employees, citizens and structures, educators, and learners. Data protection issues have always been key to education providers, but the proliferation of online learning forms and formats poses new and unique challenges in this regard. When introducing a new technology that involves the collection of sensitive data, the General Data Protection Regulation (GDPR) of the European Parliament and the Council of the European Union requires the identification and mitigation of all risks that could lead to the misuse of personal data. The article discusses some critical points regarding the application of GDPR in online learning. The goal of this article is to investigate the vulnerabilities to personal data security during online learning and to identify methods that schools and universities may apply to ensure that personal data are kept private while students utilize online platforms to learn. For the purposes of the research, the published privacy, and data protection policies of all Bulgarian universities as well as papers on how universities could adapt to the new EU General Data Protection Regulation were revised and analysed. Best practices of some foreign universities in this regard were studied as well.

Keywords: Personal data, processing personal data; GDPR; online learning; online fraud detection

1. Introduction

Information technology and the Internet made collection of personal data much easier, which can lead to harassment, identity theft, or aiding and abetting the planning of criminal acts. In online learning, data is produced through the interaction between teachers, students, and platforms. By this reason the issues of data protection and confidentiality and literacy of the participants in the personal data protection are raised. The confidentiality and protection of personal data are closely linked. Confidentiality is related to authorized access to data and information with focus on implementing policies ensuring that users' personal information is collected, shared, and used in appropriate ways. Laws, regulations, and policy documents

have been formulated at organizational, national, and international levels to protect data as an aspect of confidentiality. As distance learning is an option in the schools and universities, staff, teachers/lecturers, and students need to be informed about the importance of data protection as well as to be provided with practices that they can apply to keep them protected.

The aim of this paper is to draw attention to the risk assessment of personal data security in the implementation of the GDPR in Bulgarian schools and universities. The first part introduces the GDPR related concepts, including the idea of personal data, its life cycle, several data models for managing personal data based on its unique qualities, and data protection principles. The second part focuses on the study of threats to personal data security and outlines the measures used by schools/higher education institutions to maintain the confidentiality of personal data in online learning.

2. Personal data and their lifecycle

Any information relating to an identified or identifiable living natural person, as well as individual data that, when combined, may lead to the identification of a specific person, is considered "personal data". To clarify the methods of personal data protection, it is necessary to consider the issue of their classification. Personal data encompass many types of data that can measure and describe various aspects of an individual's identity, characteristics, and behavior. There are different classifications of personal data. Summarizing the methods adopted for the classification of personal data, (R.H. Huang at all 2020) presents a classification of 12 categories, which are presented in Table 1.

Table 1. Categories of personal data

Basic information	Name, age, place of birth, date of birth, gender, gender identity, preferences, proclivities, personal photos, race, colour, national or ethnic origin
Identification	Government-issued identification, driver's license, passport, health IDs, Social Insurance Numbers, Social Security Numbers, PIN numbers
Biometrics	Genes, fingerprints, voice prints, palm prints, auricles, irises, facial features
Authenticating	Passwords, PIN, system account, IP address, email address, security answer, personal digital certificates
Medical and Health	Physical and mental health, drug test results, disabilities, family or individual health history, health records, blood type, DNA code, medical history, medical device logs, prescriptions, and health insurance coverage
Professional	Job titles, salary, work history, school attended, education history, employee files, employment history, evaluations, references, interviews, employer data, certifications, disciplinary actions
Financial	Cars, houses, apartments, personal possessions, purchases, sales, credit, income, loan records, transactions, taxes, purchases and spending habits, credit records, credit scores, credit standing, credit capacity, physical assets, and virtual goods

Communication	Telephone recordings, voice mail, emails, SMS, phone calls, IM and social, network post, physical address, telephone number
Contact	Contact lists, friends, connections, acquaintances, associations, group membership, email address
Browsing history	Media produced, consumed, and shared: in-text, audio, photo, video, and other forms of media; Real-world and online context, activity, interests, and behaviour: records of location, time, clicks, searches, browser histories and calendar data, purchases activity, online shopping, social network profile information and the like
Device	Hardware serial number, software list, IP address, Mac address, browser fingerprint
Location	Country, GPS coordinates, room number, longitude, and latitude

3. Regulations under international frameworks for personal data protection

3.1. Models for personal data management – taxonomy

Although personal data regulations vary considerably from country to country, it is still possible to identify three main approaches: one model based on open transfers and data processing, a second model based on conditional transfers and processing, and a third model based on limited transfers and processing (M. F. Ferracane, E. van der Marel 2021). These three data models have become a benchmark for many other countries in setting their rules for both cross-border transfers and internal processing of personal data. The main features of the existing data models are systematized in the table below (fig. 1).

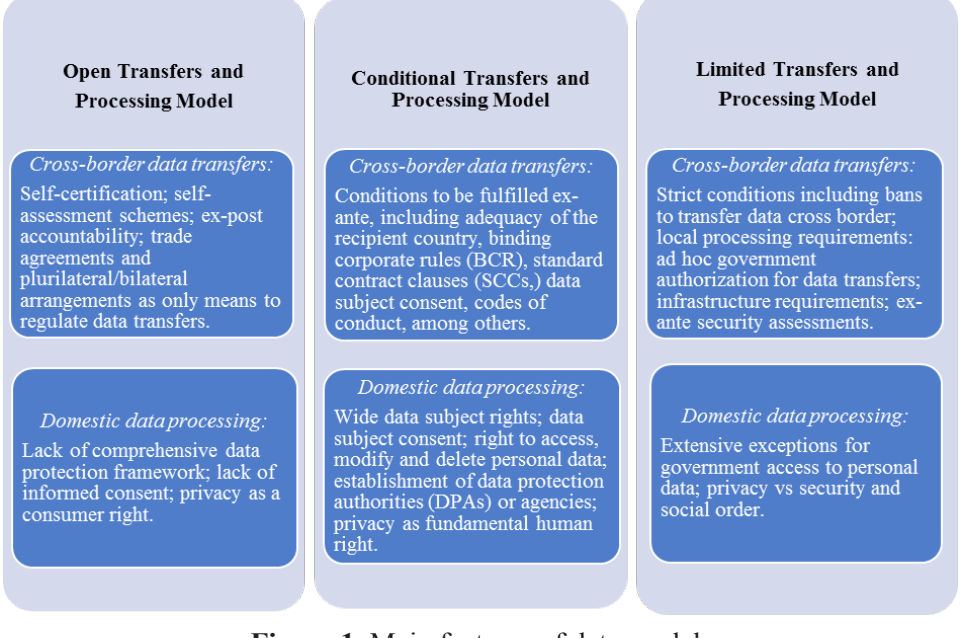

Figure 1. Main features of data models

3.2. *The personal data processing principles under the GDPR*

Personal data is processed in a variety of ways, including collection, recording, categorization, structuring, and storage. The General Data Protection Regulation (GDPR) of the European Parliament and the Council of the European Union, which was adopted on April 27, 2016, and took effect on May 25, 2018 (European Union 2016), governs the processing, storage, and use of personal data by third parties, such as individuals, businesses, and organizations. The GDPR supersedes the previous Data Protection Act (DPA) and strives to bring about a cultural shift in how we handle people's data in the digital era. It champions data subjects' rights by utilizing a variety of mechanisms, including recipient adequacy judgements, model agreements, and binding business obligations.

Seven main principles for processing personal data are specified in GDPR Article 5, which are presented in figure 2.

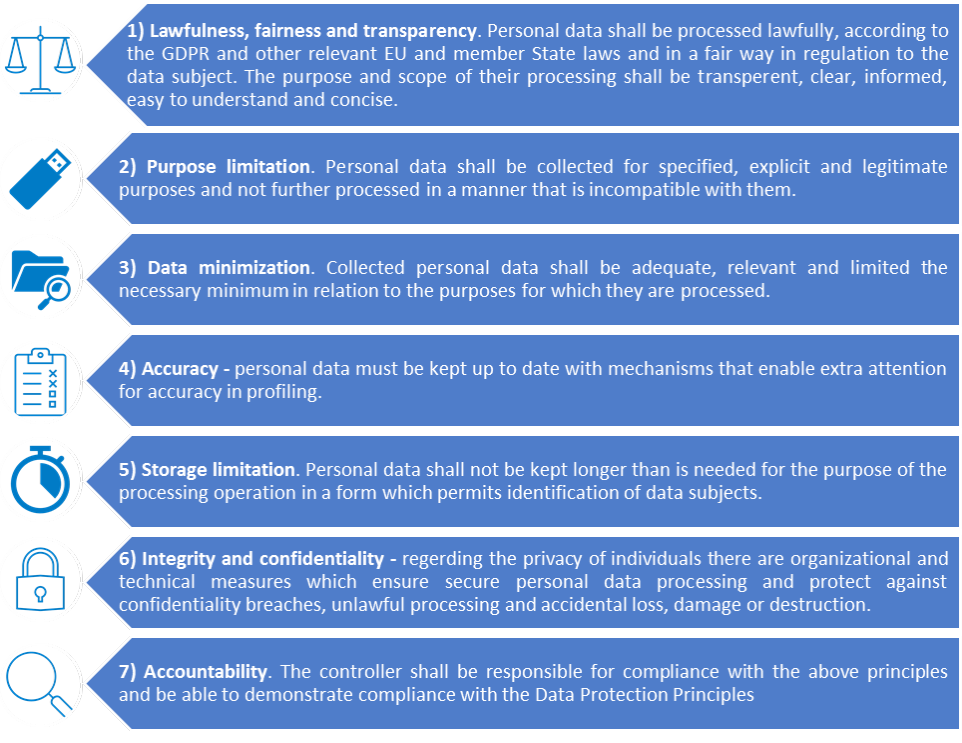

Figure 2. Principles for personal data processing

Everyone, as a user of the Internet, is a data subject. The GDPR acknowledges a slew of new privacy rights for data subjects, with the goal of giving people more control over the information they share with businesses. A person's rights in connection

to the collection, processing, and storage of personal data are outlined in the GDPR. These are the following: the right to be informed; the right of access; the right to rectification; the right to erasure; the right to restrict processing; the right to data portability; the right to object; rights in relation to automated processing and profiling.

3.3. *GDPR in schools and higher education*

Any information concerning a student's identification, academics, medical issues, or anything else that is specific to that individual student and is collected, kept, and shared by schools/ universities or technology providers on behalf of the schools/ universities is considered student personal information. Name, address, contact information, date of birth, identity documents, assessment results, student curricular record, ethnic background, language, exclusion information, and attendance information are all included. It also includes data created or generated by students or teachers/professors using technology – email accounts, online bulletin boards, work completed using an educational application or app, and anything else created or generated by or about an individual student in a learning environment.

The process of the GDPR implementation in the schools could be considered in the following six steps:

1. Creating a framework for accountability and governance. In this phase, a unit/body is set up to monitor the implementation and compliance of the GDPR by the various groups.
2. This unit/body shall specify all the criteria and requirements for compliance with the GDPR that the various departments and administrative units and services must observe and comply with in carrying out their day-to-day operations.
3. Description of the data flows, including where and how it is kept and processed, who sees it, how much is exchanged, and how it is transported, should be provided as well.
4. Evaluation of the risks such as, i.e., the probability of data being leaked, lost, erased, or stolen while passing through the school/university system.
5. A gap analysis should be done to see if the data is encrypted or stored in plain text and whether the established data flows are in line with the specified criteria, requirements and plans for ensuring GDPR compliance.

These steps should be carried out cyclically over a period and allow for the timely elimination of identified weaknesses, as well as continuous improvement and optimization of the process.

Under new GDPR rules, higher education institutions need to have organized records of what personal data exists, as well as documentation explaining why it has been held, how it was collated, who has access to it and when it will be removed or anonymised. Its need to follow several data privacy and data security requirements, such as (A. Šidlauskas, T. Limba 2019):

– Ensuring data security practices are in place.
– Implementing privacy restrictions and personal data usage policies.
– Developing a personal data consent collection process.
– Identifying a data protection officer.
– Implementing appropriate measures to protect personal data.
– Adhering to the GDPR breach notification processes.

A. Šidlauskas and T. Limba (A. Šidlauskas, T. Limba 2019) offer nine stages for implementing GDPR in higher education in 2019:

1. Get your GDPR project ready to go. An examination of the data.
2. Create a Personal Data Policy and other high-level papers; update your institution's privacy policy.
3. Make a list of all the actions that need to be completed. Validation of data.
4. Define a strategy for managing data subject rights, consent, and consents to send unsolicited, direct marketing communications, and data access, integrity, and deletion processes.
5. Conduct a risk assessment for data privacy.
6. Confidentiality in the sharing of personal information.
7. Update your institution's data processor agreements by amending third-party contracts.
8. Protect personal and sensitive data by implementing security measures.
9. Define how to deal with data breaches; security breaches should be reported to the appropriate supervisory authority.

Every educational institution in Bulgaria has developed a *Confidentiality and data protection policy*, which clearly and in detail states what personal data are collected and from whom; how the data are stored and what they are used for; who are the employees involved in data collection, storage, and management processes as well as the responsibilities associated with these activities; and procedures for documenting violations related to personal data protection and security.

4. Online learning and personal data

Data protection in online learning is a direct function of the typical steps for using multi-type online learning tools given in Table 2 (R.H. Huang et al. 2020).

Table 2. Typical steps for online learning

1) Preparing the devices, network, and tools	– Set up your device – Manage network connection on your device – Select and install learning tools – Browse the privacy policy

2) Preserving privacy when signing up/ in on learning platforms	– When registering on the platform, use a strong password to create an account – Remember username / password
3) Protecting privacy when navigating learning platforms	– Enrolling in an online course – Utilizing personalized learning services – Using search services carefully – Recognizing location services – Backing up your data
4) Staying safe while learning with social networking	– Using video conference tools with caution – Posting in the discussions and forums responsibly – Surfing the Internet safely
5) Clearing personal data after finishing learning online	– Removing data traces in online learning – Deactivating your account

By registering in the e-learning platform the following personal data of the student will be processed: user account, last name and first name, email address at home school/university or email address of the guest account, registration file data (at what time you have access to which parts of the courses), user data (content, contribution, and activities of the user in the e-learning platform). The data in the user profiles (registration data) is stored until the user profile is deleted. Course participation data (user data) is stored until the course is deleted. The data from the registration files are deleted after the end of the process of use unless the legal provisions require longer storage. Students can delete user data and voluntary entries in their user profile at any time. Recipients of the students' data are:

1. Only the staff responsible for the administration and management of the e-learning platform. They shall have access to all data stored in the system, including the log files, and may process this data solely to ensure the operation of the platform.
2. Professors shall only have access to the activities, contributions and data provided by their course participants.

Schools/universities shall provide appropriate technical and organizational measures to prohibit access to registration files or other data from the platform from which individual user profiles may be extracted. The following personal information will be processed by the school/university in online exams: name; student number; e-mail address; IP address of computer network to which computer is con-

nected; image of student card (or identity card); screen recordings of what students see on their monitors; data about website visits during the exam; webcam and audio recording of students and the room in which they are sitting during the exam. There are available different systems that provide possibility for online proctoring but all of them provide several fundamental functionalities such as detecting and disabling computer functionalities as copy-paste as well as downloading, taking images of and recording both student and screen. Another important functionality is related to analyzing the gathered data to signal irregularities that may show fraud (Aarts et al. 2021).

A classification of methods used for online exams was proposed by O. L. Holden in 2021 (O. L. Holden 2021).

Online fraud detection. Exam monitoring is sometimes known as proctoring. Exams that are proctored are exams that you take while the proctoring software watches your computer's desktop as well as video and audio from the webcam. The data collected by the control program are obtained for verification and evaluation. Video summation, web video recording, and live online proctoring are all common methods of online proctoring. Sensitive data collected when applying the method are name, video image (Basic information), irises, facial features (Biometrics), system account, email address (Authenticating). In this case, the teacher / professor should carefully review the privacy regulations and policies of the institution for access and storage. There are several approaches for implementing the online fraud detection as follows:

– Video Summation. Artificial intelligence is used in video summation software to detect fraudulent occurrences that may occur during the test. During the test, students are enrolled using their own webcam. The application will flag the video for future examination by a proctor if a fraud occurrence is discovered. These systems can produce keyframes (a collection of pictures derived from a video source) or video segments collected from a video source to depict a suspected fraud incident for human proctor identification in the future.

– Web Video Recording. In this case, the student is recorded on video throughout the exam for later viewing by the lecturer. Unlike video aggregation programs, web video programs do not have specific proctors who review all tagged instances and instead rely on review by the administrators and instructors themselves.

– Live Online Proctoring. The latest type of online proctoring uses the student's webcam and microphone to allow a live proctor to observe students during an online exam. In case the school / university chooses to use online proctoring for online exams, the educational institution thus checks which student is taking the exam and can establish that the exam rules have been followed and that no fraud was committed during the exam. Many lecturers prefer to use this type of service as it is closest to a personal exam.

Table 3 provides information on compliance with the GDPR when using different approaches to detect online fraud during an exam.

Table 3. Online fraud detection approaches

Methods	GDPR implementation
Video Summation	Images can also be considered personal data. Video records may reveal individual aspects of the students which are specially protected under the law such as race, gender, religion, and health status.
Web Video Recording	If the image or footage has not been expressly technologically processed to contribute to the identification of an individual, it is not constituted biometric data under Article 9. Schools / universities as training institutions justify the processing of data as "necessary" for the performance of their contracts with students under Article 6 (1b) of or as "necessary" for the legitimate interests of students in the timely assessment and prevention of fraud (under of Article 6 para 1f)). Personal data shall be removed after it is confirmed that a student has not acted suspiciously (Article 5 (1e)). Any such decision should be made in a reasonable time (for example, latest 30 days after reviewing the examination). Educational institutions have to obtain the requisite consent of the candidates and it is recommendable this to be inserted as a stage of the online proctoring. Moreover, an alternative for the exam should be proposed to the candidates as well.
Live Online Proctoring	The online proctoring should facilitate the examination process in line with the instructions of the educational institution. A contract between the educational institution and the online proctoring service outlines all actions and duties. After an exam has been verified, the data trail should be as short as feasible. Educational institutions must be able to demonstrate how a particular processing activity complies with the GDPR.

Knowledge-based authentication (KBA) method. This method requires students to ask multiple-choice questions based on their personal history to gain access to the exam. These questions are randomly generated from the initial profile setup questions or third-party information when the student starts the exam, and the answers are compared to verify his or her identity. The method cannot be used to monitor students' behavior during the exam. Sensitive data collected when applying the method are video image (Basic information), system account, email address (Authenticating), education history, evaluations (Professional). To catch up

with technological developments and the prevalence a data driven business models, a new General Data Protection Regulation (GDPR) will enter into force in May 2022. The compliance should be demonstrated through detailed documentation of all steps that would lead to a lawful data processing. A privacy impact assessment is required component and the results of this assessment should be taken into account from the very beginning of the method designing.

Biometrics. Biometric data that do not need physical contact with a scanner, such as height, weight, age, gender, eye color, and ethnicity, or biometric traits that do require direct physical contact with a scanner, such as a fingerprint, are another means of verification. To identify a student, this approach compares a previously registered biometric sample with freshly recorded biometric data. Biometric features should be consistent and unchanging, and the technique for gathering them should be inconspicuous and carried out by instruments that need little or no interaction. Because integrating two or more traits enhances recognition accuracy, multimodal biometric systems employ numerous biometric features and technologies at the same time to validate the user's identity.

Assess the security risks of personal data protection

An important point of GDPR is a risk assessment. In 2016, the European Union Agency for Cybersecurity – ENISA (ENISA 2016) published a set of guidelines for organizations acting as data controllers or processors to help them assess security risks and take personal data protection measures accordingly. The proposed risk assessment process is implemented in four steps:

– Definition of the processing operation and its context.
– Understanding and evaluation of impact.
– Definition of possible threats and evaluation of their likelihood.
– Risk assessment by combining the probabilities of threats and impacts.

The first step of this process defines the data processing operation in the context of risk assessment with the following set of questions:

1. What is the personal data processing operation?
2. What are the types of personal data that are processed?
3. What is the purpose of the processing?
4. What are the means used to process personal data?
5. Where is the processing of personal data carried out?
6. What are the categories of data subjects?
7. Who are the recipients of the data?

In *the second step*, the data controller must assess the impact of the loss of confidentiality, integrity, and availability. Four levels of impact are considered:

– Low – Individuals may encounter several minor inconveniences that they will overcome without difficulty;

– Medium – Individuals may face significant inconveniences that they will be able to overcome despite some difficulties;

– High – Individuals may face significant consequences that they must be able to overcome, albeit with serious difficulties;

– Very high – Persons who may face significant or even irreversible consequences that they cannot overcome.

In *the third step*, the data controller identifies possible threats and assesses their likelihood, using a set of questions that address four main dimensions of this environment:

– Network and technical resources (hardware and software);

– Processes / procedures related to the data processing operation;

– Different countries and people involved in the processing operation;

– Business sector and scale of processing.

Developed questionnaires are presented in (ENISA 2017), which organizations can use in this step.

Through this approach, the level of probability of occurrence of a threat can be determined for each of the areas of assessment as:

– Low: the threat is unlikely to materialize;

– Medium: there is a reasonable chance that the threat will materialize;

– High: the threat is likely to materialize.

The final risk assessment is determined in *the fourth step* summarizing the result of steps two and three.

After assessing the level of risk, in *step five* the organization can proceed to select appropriate security measures to protect personal data. The ENISA guidelines address two categories of measures: organizational and technical and provide a list of proposed risk level measures.

Based on the 2016 guidelines, in 2017 ENISA prepares the reports that focuses mainly on the electronic processing of personal data by organizations, which is based on IT networks and systems, as well as new digital technologies (e.g., cloud computing, mobile devices, etc.). Each of they present an assessment of security risks in the processing of personal data by a schools and university that offers a platform for e-learning and course management hosted internally on a web server. The summarized results of the four main steps of the evaluation process in these reports show that the overall risk for these two specific cases is considered medium. This methodology can be successfully used by schools and universities to self-assess risks and adopt security measures in accordance with the GDPR.

Considerations and recommendations to ensure that online training meets the requirements for privacy and personal data protection

To ensure that online learning meets the requirements for personal data protection and learners' privacy rights the following recommendations could be made:

1. At the institutional level

1.1. Educational institutions need to identify the legal basis for the processing of personal data when using online learning platforms. The legal basis for the processing of personal data when conducting online lessons or exams may be the provision of the educational service itself.

1.2. When choosing or adapting an online learning software platform, it is necessary to ensure compliance with the country's privacy laws and to ensure that no more personal data is collected than necessary. The educational institution may require students, teachers and staff to use only the institution's platform as well as institutional email addresses.

1.3 It is recommended that the educational institution conduct a risk assessment annually, which will help it to evaluate and mitigate the various risks associated with the conducting of online sessions – live streaming and/or recording. The procedure proposed by ENISA can be used for the purpose.

1.4 Rules and regulations for working in the institutional online platform should be elaborated and published on the website of the institution. Teaching and administrative staff should be trained to act in compliance with these rules.

1.5 Rules for storage, access, control and preservation of the virtual educational activities' records should be regulated.

2. At the level of teaching staff

2.1. At the beginning of each online course, learners should be informed that they are obliged to strictly follow the rules introduced by the institution for working in an online environment.

2.2. The explicit consent of all participants for recording virtual educational activity as a part of the online course is mandatory.

2.3. The written consent of the trainees to keep their cameras and microphones included during the examination procedure included in the online course must be obtained through a standard form prepared by the educational institution.

2.4. The results from the conducted online examinations should be communicated by the lecturer to the concrete student personally in a form of bilateral communication.

Conclusion

Successful implementation of the new GDPR rules in the educational institutions requires balance of system, process and privacy resources, as well as proper methodology performed by a team. The article presents the main steps of the process for the application of this regulation, the main points of its application in the various methods of online exams, as well as a four-step method for risk assessment.

Acknowledgement

In this paper are presented some results obtained in the framework of more comprehensive research conducted under the Erasmus+ project EDucational University GATeway to enhance innovative E-learning capabilities, resilience, and new best practices (EDU-GATE) № 2020-1-IT02-KA226-HE-095538.

REFERENCES

AARTS, E., FLEUREN, H., SITSKOORN, M., WILTHAGEN, T., 2021. The New Common (How the COVID-19 Pandemic is Transforming Society). Springer, Cham. https://doi.org/10.1007/978-3-030-65355-2.

Edu-Gate: ENISA, Guidelines for SMEs on the security of personal data processing, 2016, https://www.enisa.europa.eu/publications/guidelines-for-smes-on-the-security-of-personal-data-processing.

ENISA, Handbook on Security of Personal Data Processing, 2017, https://www.enisa.europa.eu/publications/handbook-on-security-of-personal-data-processing. European Union, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (GDPR), O.J. (L 119) 32, European Union, Brussels, 2016, http://data.europa.eu/eli/reg/2016/679/oj.

FERRACANE, M. F. E van der Marel, Regulating Personal Data: Data Models and Digital Services Trade, World Bank Policy Research Working Paper No. 9596, World Development Report 2021, March 2021, https://openknowledge.worldbank.org/bitstream/handle/10986/35308/Regulating-Personal-Data-Data-Models-and-Digital-Services-Trade.pdf.

HOLDEN O. L., M. E. NORRIS, V. A. Kuhlmeier, Academic Integrity in Online Assessment: A Research Review, 14 July 2021, https://www.frontiersin.org/articles/10.3389/feduc.2021.639814/full.

HUANG R.H., LIU, D.J., ZHU, L.X., CHEN, H.Y., YANG, J.F., TLILI, A., FANG, H.G., WANG, S.F., 2020. Personal Data and Privacy Protection in Online Learning: Guidance for Students, Teachers and Parents. Beijing: Smart Learning Institute of Beijing Normal University.

SWAMINATHAN N. What Is Columbia Doing With Your Data?, THE EYE | FEATURES, March 01, 2019, https://www.columbiaspectator.com/the-eye/2019/03/01/what-is-columbia-doing-with-your-data/.

SHAW M., M. HALKILAHTI, M. REISSMAN, S. RUIZ RUIZ, J. VOUTILAINEN. Futures Personal Data Can Build: Scenarios for 2030, In book: COOLEST STUDENT PAPERS AT FINLAND

FUTURES RESEARCH CENTRE 2017 – 2018, Publisher: Finland Futures Research Centre University of Turku, 2018,
https://www.researchgate.net/publication/336923184_Futures_Personal_Data_Can_Build_Scenarios_for_2030.

ŠIDLAUSKAS A., T. LIMBA. General data protection regulation implementation in higher education institutions, Proceedings of EDULEARN19 Conference 1st-3rd July 2019, Palma, Mallorca, Spain, 2040 – 2047.

✉ **Dr. Evgeniya Nikolova, Assoc. Prof.**
ORCID ID: 0000-0001-8313-1572

✉ **Dr. Mariya Monova-Zheleva, Assoc. Prof.**
ORCID ID: 0000-0001-8910-2502

✉ **Dr. Yanislav Zhelev, Assoc. Prof.**
ORCID ID: 0000-0003-2783-5617

Institute of Mathematics and Informatics Bulgarian Academy of Sciences
Laboratory of Digitization – Burgas
Burgas Free University
5, Demokratsiya Blvd.
8000 Burgas, Bulgaria
E-mail: evgeniyanikolova@gmail.com
E-mail: mariya@zhelev.com
E-mail: yanislav@zhelev.com