

## LMS DATA COLLECTION, PROCESSING AND COMPLIANCE WITH EU GDPR

A. Aleksieva-Petrova<sup>1</sup>, I. Chenchev<sup>2</sup> and M. Petrov<sup>3</sup><sup>1</sup>Technical University of Sofia, Faculty of Computer System and Technology (BULGARIA)<sup>2</sup>Technical University of Sofia, Faculty of Computer System and Technology (BULGARIA)<sup>3</sup>Sofia University "St. Kliment Ohridski", Faculty of Mathematics and Informatics (BULGARIA)

## INTRODUCTION

Learning Management Systems (LMS) are enterprise-wide and internet-based systems, which integrate a wide range of pedagogical and course administration tools. Authors in [2] are talking about a significant change, that takes place in higher education. It is related to the integrated computer systems known as LMS. They allow instructors and students to share course materials, make class announcements, submit and return course assignments, and communicate with each other online [6]. A detailed overview of the rapid evolution of online LMS and a comprehensive review of their influence on the teaching and learning is presented in this paper. We are sharing their concerns about possible corporatization of academic knowledge and the importance of steps that have to be taken to identify how LMS systems can be used to augment and complement an institution's core teaching objectives. Therefore, the collecting of the statistical and analytical data becomes very important.

One of the most widely used LMS is the Moodle system [5]. It processes few types of data – the online data like data for users, courses, assignments, results and real-time activities. The both of those types are stored in a database. Together with the general information – like course materials, a personal information is also collected and stored in the Moodle's database. It is also stored in the database backups, in the snapshots of the system and in the different archives.

Moodle processes different types of data stored in tables, logically grouped in two major types: information for users and courses, and activities. The latest version (at the time of paper preparation is 3.6.3+, released on date 2019-05-03) of Moodle database has 446 tables (with eleven plug-ins like VPL, Questionnaire etc.). This is an upgraded version (not new installation), because the research was performed in a life Moodle system. Only thirty-one of them contain the word 'user' in their name and another fifteen tables have the word 'log' in their names. There already exist some plug-ins for Moodle's compliance with the GDPR regulation.

## LMS DATA STRUCTURE ANALYSIS

For the data structure analysis we use following system settings: Moodle LMS version is 3.6.3+ (Build: 20190503).

- Structure of Moodle System - Database
- Structure of Moodle System – Log levels
- Data types categorization analysis, based on Moodle LMS

## LMS SYSTEMS COMPLIANCE WITH GDPR

In order to define methodology of GDPR principles in the LMS on the fig. 1 we define a hierarchical model of the different locations, where personal data can reside for a given LMS system:

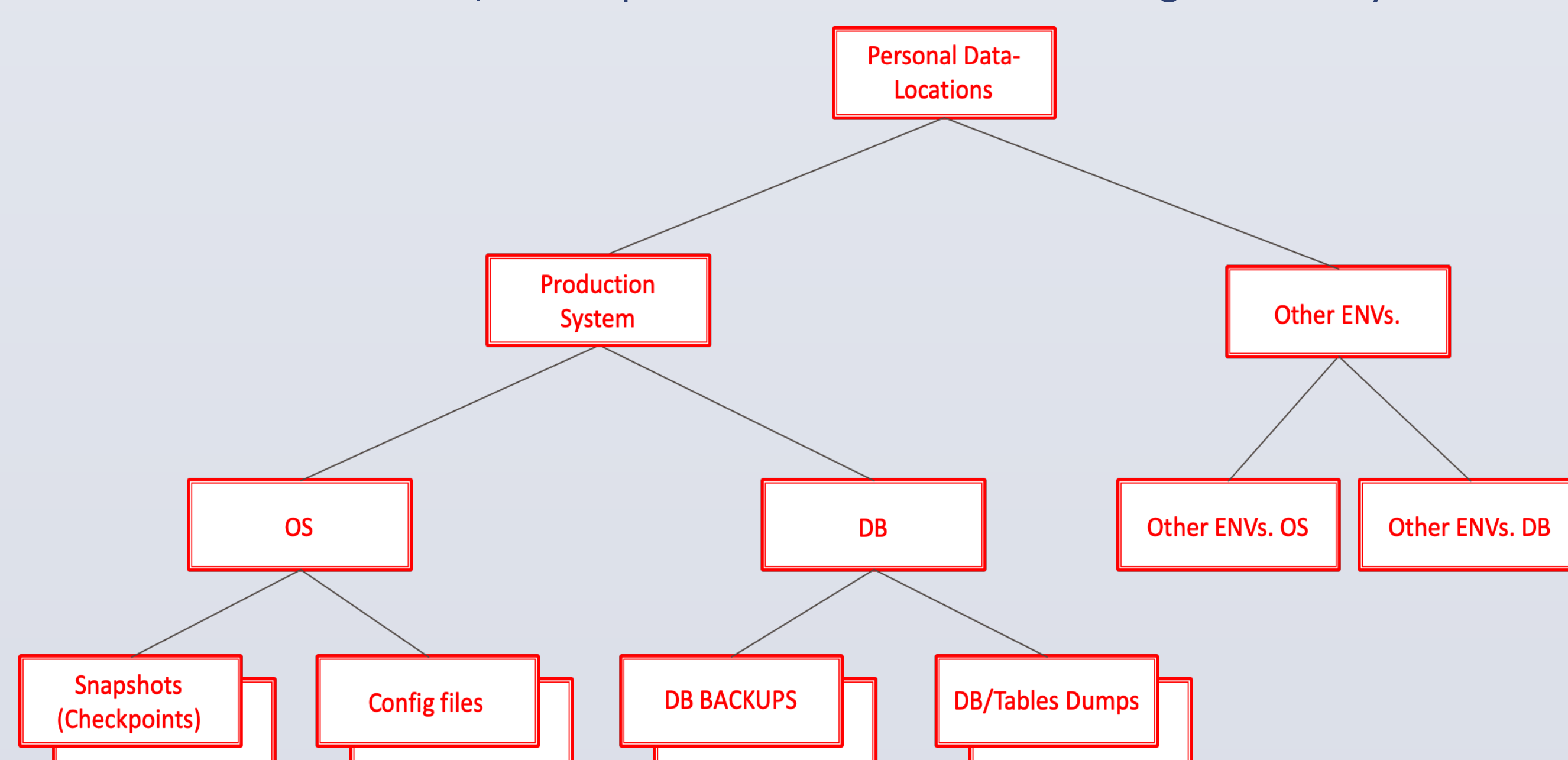


Figure 1. Personal Data Locations – LMS Live System

In a life LMS system (for the purpose of this research is the Moodle system), all of those places/locations can contain personal information. Obviously, it is not enough to develop a plug-in(s) to only manage the Moodle's database. For a LMS system, to comply with the GDPR regulation is necessary to extend the scope.

In order to define the GDPR personal data in LMS, it is necessary to make review of the answers, given to number of questions.

**The first question:** Why it is necessary to held personal information in LMS?

**The second key question:** How to store the information?

**A third major question** (relates to the problem of) Who can access to those data?

It is necessary to take into account when the personal data will be deleted or eliminated. This should be the right of every user to ask for his / her data to be deleted. Then two basic questions arise. The first sub-question is to delete the data from the archives and the logs. The second key sub-question is how this will affect the process of student learning and assessment.

As result in next section of paper we propose model of data anonymization, using results of analysis of answers of questions above.

## LMS DATA PROCESSING APPROACH

The process is divided into two sub-processes – data anonymization process – with four steps (and one pre-step), and data-anonymization sharing for processing by external tools – in two steps.

Data anonymization process (as depicted on fig.1) contains four steps: Pre-step - zero: initial data, step 1 - User data (filter by policy) – directly identifiable, User 'data' (filtered by preferences) – indirectly identifiable, step 3 - Enforce anonymization usage, and step 4 – Anonymous data – with their usage attributes (what can be done and how can be shared). In last step – data are already filtered in previous steps and contains only valid for data processing, and how these data are related for processing with (possible) external systems and tools.

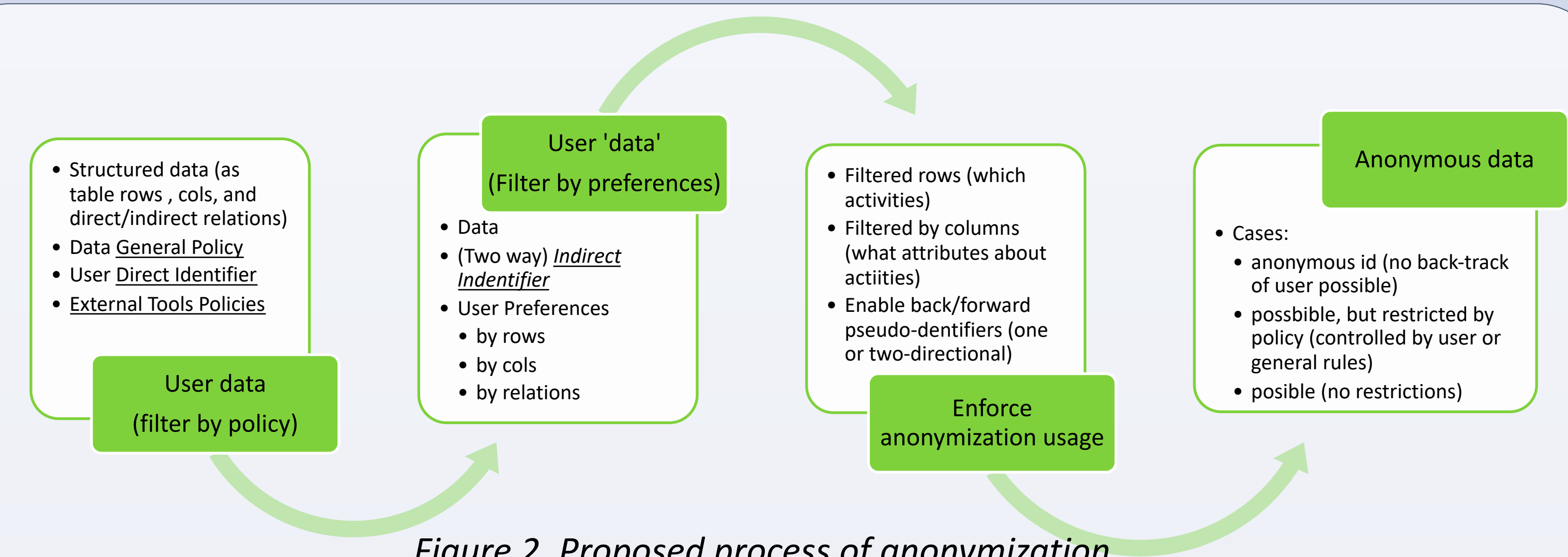


Figure 2. Proposed process of anonymization

Second sub-process is called data anonymization sharing. After completing the first process of data filtering and tagging with shared attributes – the data can/cannot be shared with external processors (as depicted on fig. 2). In this scenario from previous sub-process have four use cases – 1) user does not agree to share anonymous data with external tool or with specific tool - in this case we cannot apply sharing anonymous data. 2) user agrees to share anonymous data, but no back-track possible; 3) user agrees to share data and in very strictly controlled cases agrees to be possible to de-anonymize data (deanonymization for external tool or deanonymization for internal system, without giving feedback to external tool); 4) possible de-anonymization on request by external tool.

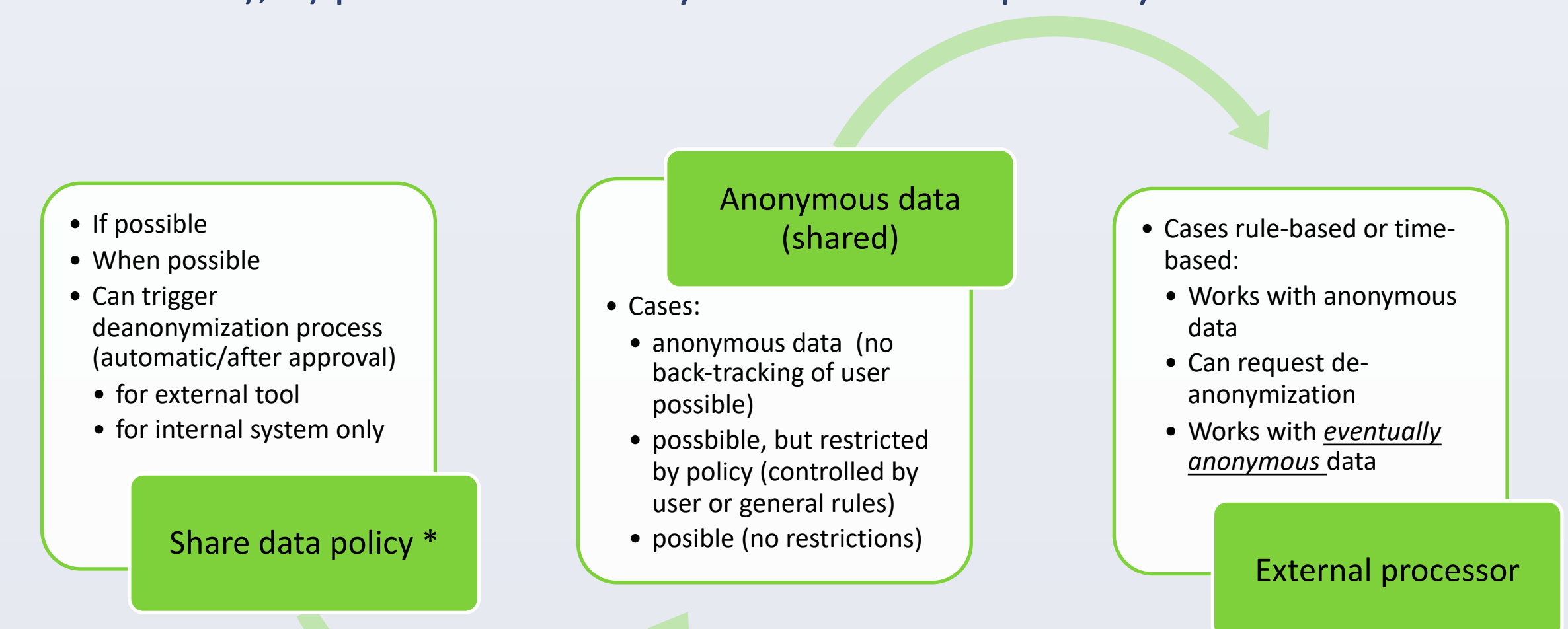


Figure 3. Process of sharing data to external systems

## CONCLUSION

Definitely, the area for Moodle's compliance with GDPR is still to be explored. The following points highlights three possible areas of further research:

1. Identification and preparation of proper backup/restore strategy for Moodle system in compliance with the GDPR regulation;
2. Preparation of methodology for automatic data anonymization - in order to collect statistical and analytical data, such as usage of the system, usage of the courses, behavior of the students in learning process of different areas, reports and etc. for further research.
3. Define a methodology for GDPR compliance in the LMS systems in general – from the initial systems architecture design, to the data collection, data processing, storing and releasing/deleting the personal data.
4. Validate proposed approach for Anonymization and LMS Data processing with real example.

Points for improvements and further research gives starting point for different directions of research, nevertheless we believe that proposed classification, models and approaches are significant step to creation of software platform for analyzing of learning and game data for a user-oriented adaptation of technology-assisted training by providing framework for anonymization. Anonymization is crucial point to system, that analyses big data, related to any kind of users.

## ACKNOWLEDGEMENTS

The research in this paper was supported by Contract KP-06-OPR/1 from 13.12.2018 project "An innovative software platform for big data learning and gaming analytics for an user-centric adaptation of technology enhanced learning (APTITUDE)", Competition for financial support of basic research projects on societal challenges – 2018, Bulgarian National Science Fund, Ministry of Education and Science, Bulgaria.

## CONTACTS



**Assoc. Prof. Adelina Aleksieva-Petrova, PhD**  
Technical University of Sofia, Bulgaria,  
Faculty of Computer Systems and Technologies  
Telephone: (+ 359) 877 706 778  
E-mail: aaleksieva@tu-sofia.bg  
adelina.aleksieva@gmail.com



**Ivaylo Chenchev, MSc**  
Technical University of Sofia, Bulgaria,  
Faculty of Computer Systems and Technologies  
Telephone: (+ 359) 888 500 509  
E-mail: ivaylo.chenchev@gmail.com



**Assoc. Prof. Milen Petrov, PhD**  
Sofia University "St. Kliment Ohridski",  
Faculty of Mathematics and Informatics,  
Department of Software Engineering, Bulgaria  
Telephone: (+359) 878 594 220  
E-mail: milenp@fmi.uni-sofia.bg  
milen.petrov@gmail.com

