Information Services & Use 41 (2021) 61–69
DOI 10.3233/ISU-210107
IOS Press

Thinking with GDPR: A guide to better system design

Andrew Cormack*
Chief Regulatory Adviser, Jisc, Lumen House, Harwell Oxford, Didcot, UK

Abstract. Europe's General Data Protection Regulation (GDPR) has a fearsome reputation as "the law that can fine you €20 million." But behind that scary slogan lies a text that can be a very helpful guide to designing data processing systems.
This paper explores that side of the GDPR: how understanding it can produce more effective - and more trustworthy - systems. Three popular myths often take designers down the wrong track: that GDPR is about stopping processing, is about users, and is about consent. Instead we consider, from a design perspective, the GDPR's source material, its Principles, and its Lawful Bases for processing. Three examples - from the field of education, but widely applicable - show how "thinking with GDPR" has improved both the effectiveness and safety of large-scale data processing systems.

Keywords: System design, information assets, data quality, data protection, general data protection regulation, GDPR

What is the GDPR

Myth: "It's all about stopping …"

Formally, the General Data Protection Regulation (GDPR) is Regulation 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data [1]. Its very title contains two, balanced, purposes: "protection of natural persons" and "free movement of [personal] data." Rather than a law that prohibits, it is better viewed as a law that enables, subject to safeguards. That becomes clear if you approach it expecting binary answers to 'prohibit' questions: "can we do X?" The GDPR's most common response: "yes, provided Y." Only occasionally - such as the content of privacy notices and data processor contracts - does it prescribe what must be done. More often it sets pre-conditions and outcomes. Considering it first as a specification document, rather than a law, may therefore be fruitful.

The legally-binding GDPR comprises ninety-nine articles: about half address definitions and enforcement so are not directly relevant to system designers. Much can be learned from just two - Article 5 on Principles and Article 6 on Lawfulness - examined in the next two sections. There are also one hundred and seventy-three Recitals, which set out legislators' thinking, but are not formally binding in law. These also contain useful design guidance, for example Recital 71 on automated decision-making is highly relevant to algorithms and Machine Learning.

*Tel.: +44 01235 822200; E-mail: Andrew.Cormack@jisc.ac.uk.

Regulators and courts apply the GDPR to specific situations, markets and technologies:

- Collectively - as the European Data Protection Board (EDPB) - the gathering of national data protection regulators publishes Opinions, for example on video devices, connected vehicles and social media [2]. Before 2018 the group was known as the Article 29 Working Party and many older opinions are still relevant [3];
- Occasionally, the actions of legislators or regulators may be reviewed by the European Court of Justice, which is the ultimate authority on the interpretation of European law [4];
- Individual regulators have published valuable guidance including the European Data Protection Supervisor - effectively the "national" regulator for the EU institution - on scientific research [5,6], France's CNIL on voice assistants [7], and the UK Information Commissioner on Artificial Intelligence [8].

Thinking with principles

Myth: "It's all about users …"

Although the GDPR is, ultimately, about protecting individuals, its duties fall almost entirely on the organisations - which the Regulation calls "Data Controllers" - that decide why and how to collect, store, use, share and dispose of (or anything else within the broad definition of "processing") data. In particular, these entities must ensure compliance with seven Principles, set out in Article 5, which are key requirements for all system designers.

Lawfulness, fairness and transparency. Lawfulness within GDPR is discussed in the next section, but processing must also be lawful under other laws. Fairness and transparency are linked, in that transparent activities are more likely to be fair. However, fairness also relates to individuals' and society's expectations: simply declaring unexpected processing or purposes in your privacy notice does not "whitewash" their unfairness [9].

Purpose limitation. This links to transparency: individuals must be told what the purpose(s) of processing are. But it is also a safeguard. Organisations must have clear purpose(s) in mind when they collect personal data and beware of that purpose subsequently shifting. Limited extensions may be covered by the "compatible purposes" concept (Article 6(4)). But if a new purpose is not compatible, each individual must consent to reusing existing data. Clear purposes at the start of any design will avoid a lot of remedial effort.

Data minimisation and **storage limitation**. These require, respectively, that personal data are only collected or processed if needed for the purpose, and that they are kept for no longer than the purpose requires. Data minimisation encourages pseudonymisation: keeping information that links an individual to their record either separately or not at all. Storage limitation encourages data reduction: anonymising/pruning older data, then summarising them in statistics. The GDPR's broad view of "personal data" highlights that pseudonyms and rich data records still represent risks, even without names and identification numbers. Even "anonymous" data need a continuing process to monitor the risk of de-anonymisation [10]. Well-designed retention periods - thinking "how soon can I delete it?" rather than "how long should I keep it?" - can reduce that effort, as well as the impact of security breaches.

Accuracy. This is a positive obligation on data controllers: they should not rely on individuals providing corrections (Article 16) and they must ensure that the data remain up-to-date. Accurate data are often a functional requirement anyway, but this Principle reminds designers to be realistic about data durability.

Again, this links to storage limitation: before old data are likely to become inaccurate, either refresh them, or reduce them to a historic trend statistic.

Integrity and confidentiality. This requires organisations to provide appropriate protection for the data that they hold and process. They should consider technical and organisational measures, both preventive and reactive, against malicious and accidental acts by outsiders and insiders: good security requires multiple components playing complementary roles. These should reflect the sensitivity of the particular processing activity and data, and the current state of the arts of defence and attack [11].

Accountability. This is an over-arching obligation on organisations to be able to demonstrate how processing satisfies, and will continue to satisfy, the Principles. It is not primarily about enforcement or blame - although lack of Accountability leads to difficult conversations with Regulators - rather it is about acting responsibly. For designers, this means the Principles should be relevant from initial design, though implementation, operation and disposal. Accountability means focussing on risks to individuals, not the organisation, using tools such as Data Protection Impact Assessments (DPIAs) [12] and Purpose Compatibility to examine systems, processes, safeguards (against both error and misuse), audit and review. Transparent Accountability shows data subjects and stakeholders that the organisation understands what it is doing, and builds trust that its systems will respect individuals and their data.

Thinking with lawful basis

Myth: "It's all about consent …"

Article 6 expands the Lawfulness Principle: processing must meet one of six conditions, commonly referred to as the "lawful bases". Five start "processing is necessary for …" and cover direct consequences of a situation (e.g. a health emergency) or decision an individual has taken (e.g. to enter employment). The term "necessary" is a safeguard with specific meaning: that the purpose "cannot be achieved by other means" (Recital 39). Consent is the sixth basis: potentially covering processing that is not "necessary", and best suited for adding information to an existing relationship.

The six, and the kinds of guidance that they provide, form three pairs. Two - delivery of **contracts** and protecting **life** - were identified by the original legislators [13]; two let future legislators create mandatory **legal obligations** and permissive **public tasks**; two, without further legislation, recognise **legitimate interests** of organisations and **consent** by individuals. The pairs produce different types of safeguards for individuals and guidance for designers.

For the first pair - contract (Article 6(1)(b)) and vital interests (Article 6(1)(d)) - the GDPR is the only opportunity to provide legal safeguards. Both are therefore defined narrowly. To be "necessary for contract", processing must relate directly to the substance of the agreement not, for example, secondary fund-raising [14]. Vital interests only covers imminent threats to life or serious injury. These bases help system designers distinguish the core data and processing needed to achieve a purpose from supplementary actions that may be part of the same transaction, but need their own lawful basis [15].

For legal obligation (Article 6(1)(c)) and public task (Article 6(1)(e)), safeguards were left to later national (or European) legislators. The GDPR drafters merely outlined - in Articles 6(2) and 6(3) - the kind of legislation they expected: it should be "Union or Member State law", though "not necessarily … a legislative act" (Recital 41); it should "determine more precisely specific requirements"; and it should "meet an objective of public interest and be proportionate to the legitimate aim pursued". Ideally, especially from a designer perspective, it should specify "types of data, … data subjects concerned, …

entities to and purposes for which personal data may be disclosed, … purpose limitation, … storage periods, … processing operations and processing procedures" (Article 6(3)). In practice, few do, though legal obligations are typically better defined than public tasks.

Finally, legitimate interest (Article 6(1)(f)) and consent (Article 6(1)(a)) had to allow a broad range of processing - including for purposes and by means not envisaged in the late twentieth century - without future legislative safeguards. The GDPR therefore specifies - very helpfully for system designers - the processes organisations must use to determine whether a purpose and means of processing is appropriate.

Legitimate interest - where the organisation makes the ultimate decision to process - has three layers of safeguard: first the purpose of processing must itself be legitimate which, according to Recital 47, depends on the individual's relationship with the organisation and whether they would have "reasonably expect[ed]" the processing when their data were collected. Two specific legitimate interests - information security and administrative functions within a corporate group - are discussed in Recitals 49 and 48. Second, as above, the processing must be "necessary" to achieve the purpose. And third, unique to this basis, the organisation must balance the impact on all the individual's rights, freedoms, and interests (not just privacy) against the organisation's interest [16]. Even legitimate processing may fail this "balancing test", making it both a strong safeguard and, if explained clearly and publicly, a confidence-building measure for those whose data are processed.

As the widest lawful basis, consent has the tightest procedural safeguards, set out in binding Articles, not just Recitals. Although individuals have the ultimate choice whether or not to consent, organisations (and designers) must work hard to establish a context where their choices are lawful. Consent must be free, so is only appropriate for decisions "without detriment" (Recital 42) to the individual. In particular, consent must not be a condition of providing a service (Article 7(4)); it is presumed invalid where the parties have a significant power imbalance, for example employer and employee (Recital 43). More generally, this casts doubt whether consent can be used where one of the "necessary" bases applies, since refusing consent will result in a detrimental loss of opportunity. Consent must be signalled by a positive act (Recital 32), fully informed, with the consequences of both granting and refusing set out in "clear and plain language" (Article 7(2)). Thus only decisions with simple, predictable, consequences are likely to be suitable for consent. The individual must be able to withdraw their consent - halting further processing - at any time and as easily as they granted it in the first place. Organisations must keep records demonstrating that consent was obtained in accordance with these and other conditions.

Finally, the safeguards provide a useful check that the appropriate lawful basis was chosen. If they seem hard to meet in a particular circumstance then consider whether another basis applies, or if the purpose or processing may actually be unlawful. They also contribute to the Accountability principle: ensuring organisations bear the burden of designing systems and processes that are safe for their customers and users, rather than passing on choices that the organisation found too hard.

GDPR thinking in practice

Voter registration

When the UK's *Higher Education and Research Act 2017* required English universities to help students register to vote (s.13(1)(f)), thoughts turned to large databases, complex controls, participation dashboards, and a tangle of legal rights and obligations. It wasn't even clear whether the law's heavily-qualified phrasing made such processing a legal obligation, a public task, or something else.

Applying a GDPR-thinking lens, purpose and minimisation revealed a much simpler core function. At an individual student's request, gather the data that they need to register and submit it to the right Electoral Registration Officer (ERO) for their residential address. This last point is the most complex, with universities in cities, in particular, having students living in many different constituencies.

The lawful basis is consent, since students can freely choose between our system, postal, or on-line registration; the data routing purpose and function is simple to explain. In terms of the Principles, we have a single purpose and lawful basis; simple documentation provides fairness and transparency. We achieve data and storage minimisation by gathering only the information needed to register and deleting it when it is forwarded to the ERO. Ideally this would happen immediately, but few EROs have an Application Programming Interface (API) for registration. Instead, having identified the appropriate ERO, we know whether they prefer transfer by API call, a batched spreadsheet, or even a printout in an envelope [17]. A short time limit for transfers reduces delays in registering and minimises storage duration. Information gathered from the student record system - which both the student and university depend on for other purposes - should be the most accurate available, reducing the checking requirement on EROs. For technical security students use the university's single-sign-on (SSO) system to gather their data; a 'register' button consents to release to the ERO. Even the anticipated audit functions are unnecessary: SSO logs show universities when (and if necessary which) students release data to the system. To measure uptake, we can provide statistics on how many actually register. Separate logs are only kept for fault-finding.

This simple result of GDPR-thinking delighted its customers.

Federated authentication and authorisation

Whereas GDPR-thinking about voter registration kept the same data flow in a much simpler technical system, federated authentication and authorisation reduces data flow using a more complex technical and organisational structure.

Students and staff needing external content, such as licensed journals, to learn and study used to create personal accounts with the content provider. These required a lot of personal data, with entitlement "proved" by knowledge of a code (often stuck on a library notice board) or an Internet Protocol address associated with their institution. Providers could see exactly who was reading what: institutions depended on them reporting which licenses were being used.

On examination, these flows have two Purposes. Service providers want assurance that content is going to members of institutions that pay for it, and that individuals who misuse content and systems can be dealt with. Both are badly served by self-asserted data. Giving institutions a greater technical and organisational role delivers both purposes much more effectively and greatly reduces the disclosure of personal data.

The technical component uses the Security Assertion Markup Language (SAML - see: https://en.wikipedia.org/wiki/Security_Assertion_Markup_Language) to let three parties - user, content provider and institution - exchange cryptographically-signed statements. The user requests access to content; the content provider forwards an authentication request to their institution. The user logs in (using their institutional SSO credentials); the institution tells the provider whether authentication succeeded. The provider then asks the institution for the information it actually needs; e.g. whether the user is a student or member of staff - to decide whether or not to grant access [18].

This technology is embedded within an organisational layer where the institution makes two key promises: (1) that it will provide accurate information (typically, as accurate as used for internal systems); and (2) that it will apply effective disciplinary processes if authenticated users misbehave. Rather than

bilateral contracts, common "federation" agreements are normally used (e.g. [19,20]). This combination of roles benefits all three parties: service providers get the information they need for authentication and an effective solution to misuse; institutions get visibility and control of license usage (though, normally, not individual activity); and, because verified information is exchanged, far less of it is needed.

This approach aligns operational incentives with the GDPR Principles. The information provided is limited, both by design and agreement, to access control and personalisation purposes. Storage Limitation and Accuracy combine because service providers bear the risk of not requesting fresh data: many do so on every successful authentication. Data Minimisation has helped identify groups of service providers with similar data requirements. Many need only the relationship between user and institution (using a defined vocabulary) plus a unique opaque identifier for that user on that service (but not others) to store preferences and search results between sessions. Others - such as group management and discussion forums - need a verified name and email address to link online and real-world personas. Policy and technical standards to support new categories continue to be developed [21]. Integrity and Confidentiality are addressed both in federation agreements and technology.

This processing should have a common lawful basis. But, as best illustrated by researchers, sometimes a particular paper is necessary for their (contracted) employment, but sometimes it is not. We cannot rely on consent for the former, because compelled consent is invalid, nor can we rely on contract for the latter. The institution could, perhaps, identify which was which, but only by privacy-invasive scrutiny of what researchers are reading and why. Even if the institution knew, the three-party relationship would require new technology to inform the service provider.

A different lawful basis simplifies the technology and relationships: each institution and service provider has a legitimate interest in delivering the information its users wish to access. This supports the Principles by making each party distinguish information necessary to deliver its service from optional information that it can use if provided. And it provides safeguards: respectively the legitimate interests balancing test (promoting Data Minimisation) and either true consent (e.g. to add an avatar) or necessity for an agreement (e.g. receiving email updates).

Analytics

The final example shows how GDPR thinking can build trust in an activity that might otherwise be tainted by others' Big Data practices.

Society's increased digitisation has created opportunities for "analytics": using data generated during a process to improve that process. In education this has focussed on using data from Virtual Learning Environments (VLEs) to analyse the effectiveness of teaching materials ("curriculum analytics") or provide personalised help to students ("learning," or "learner analytics"). Analytics might, for example, support students struggling with a particular topic, those insufficiently engaged with their studies, or those unhealthily perfectionistic [22,23].

Viewing pilots as human subject research, ethics boards naturally relied on volunteers' consent. But once institutions build analytics into all students' education, GDPR consent is almost certainly the wrong basis. It is doubtful whether students can give free consent, for example during enrolment, to processing that will be a core part of their education; since analytics is developing rapidly, it would be hard, or impossible, to state what processing will occur during a three-year degree or even a one-year course; statistical models using opt-in data will omit the experiences of disengaged students, who are supposed to be among the main beneficiaries [24].

One option is to view improving a process as a compatible purpose to performing it, with the same lawful basis. Depending on how an institution and its national legal system treat education that might be public task or contract.

Purpose Compatibility is one safeguard, but more can be added by separating the analytics process into five stages [25]. Information is **collected or observed** during an ongoing process; self-reported information (such as reasons or sentiment) may be **donated** by those engaged in the process; **analysis** of that information identifies relevant patterns; those patterns are used to **improve** the process or offer personalised **interventions** to individuals. GDPR thinking about each stage produces richer guidance and stronger safeguards - and much clearer Accountability - than a single purpose or premature "consent" decision.

Data **collection** contributes to two purposes: identifying which sources are most informative for the statistical model, then applying that model to live data [26]. Some sources that are necessary (to consider) in the first stage should be found to be not necessary (to use) in the second. Understanding these as different Purposes results in different safeguards: source identification should - borrowing from GDPR Article 89 - use pseudonymous or anonymous data, and never result in decisions affecting individuals; model application uses less data, but may require links to individuals, the legitimate interests balancing test ensures that collection does not create inappropriate risks. Treating both as necessary, with strong safeguards, permits the use of whole-cohort data. For **analysis** of live data the balancing test is again an important safeguard, ensuring that models and processing are not unfair or discriminatory; Recital 71 on fairness in algorithms, plus the extensive literature on transparency [8] are valuable design guides. Some process **improvement** will not need personal data, but analysing students' activity data may generate data about their tutors. If so, a legitimate interest analysis of *their* rights and freedoms - including rights such as free speech - can avoid problems. Even under laws that consider education (and improving it) a public task, these balancing tests provide a valuable safeguard.

This leaves two stages - **donation** and **intervention** - where practicality requires something close to a consent process anyway. Students can, and will, refuse to donate unless they believe it is safe. When an intervention is offered (for example an invitation to a support session) they can ignore it. These stages must expect incomplete and inaccurate data: from students who don't participate, or lie. Using consent here has a further, personal and legal, benefit: the institution can seek permission for a single narrow purpose, at a time when it can fully explain the implications [27].

Finally, the GDPR Principles highlight an interaction between Accuracy and Storage Limitation. Since the purpose of analytics is to improve processes, success should make older data inaccurate, as the process that generated them will have changed. Well-designed analytics processes should consciously re-classify, and summarise, data more than a few cycles old as meaningful only for trends, not for current models.

Addressing widespread concern about "Big Data" [28], GDPR thinking provides strong guidance and safeguards, creating analytics processes that are better understood and more trustworthy.

Summary

This paper presents the GDPR in a new light: as a rich source of guidance for system and process designers. Applying two key Articles - the Principles in Article 5 and the Lawful Bases in Article 6 - reveals many ways to improve three real-world data processing activities. Perhaps surprisingly, we have not needed to discuss whether any particular value or record is personal data. The advantages of GDPR thinking - to get more benefit from data, while managing the risks of such use, thus building confidence and trust - need not be so limited.

About the Author

Andrew Cormack has worked for the UK's national research and education network for more than twenty years. He joined as Head of the Computer Emergency Response Team, but for the past fifteen years has focussed on policy and regulatory issues. His role has expanded from networks to technical services such as federated authentication and, more recently, services based on data. Throughout, his aim has been to make law and technology work together, rather than in opposition.

To support this, he has added bachelors and master's degrees in law to his original mathematics degree. He is a regular speaker at national and international conferences, and writes in all formats - from blogs to peer-reviewed papers. He chaired the programme committees for the 2009 TERENA Networking Conference and 2019 FIRST Security and Incident Response Conference. In 2015 he was awarded the Vietsch Foundation Medal for his role in advancing trust and security within the European research and education sector. E-mail: Andrew.Cormack@jisc.ac.uk; Phone: +44 01235 822200. ORCID https://orcid.org/0000-0002-8448-2881.

References

[1] *Regulation 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.*

[2] European Data Protection Board, *GDPR: Guidelines, Recommendations, Best Practices,* https://edpb.europa.eu/our-work-tools/general-guidance/gdpr-guidelines-recommendations-best-practices_en, accessed July 12, 2021.

[3] Article 29 Working Party, *Archives 1997–2016*, https://ec.europa.eu/justice/article-29/documentation/index_en.htm accessed July 12, 2021.

[4] Court of Justice of the European Union, *Fact Sheet: Protection of Personal Data,* https://curia.europa.eu/jcms/upload/docs/application/pdf/2018-10/fiche_thematique_-_donnees_personnelles_-_en.pdf, accessed July 12, 2021.

[5] European Data Protection Supervisor, *Case-Law & Guidance,* https://edps.europa.eu/data-protection/eu-institutions-dpo/case-law-guidance_en, accessed July 12, 2021.

[6] European Data Protection Supervisor, *Preliminary Opinion on data protection and scientific research*, 6 January 2020, https://edps.europa.eu/data-protection/our-work/publications/opinions/preliminary-opinion-data-protection-and-scientific_en, accessed July 12, 2021.

[7] Commission Nationale de l'Informatique et des Libertés, *"On the record": CNIL publishes a white paper on voice assistants*, 16 December 2020, https://www.cnil.fr/en/record-cnil-publishes-white-paper-voice-assistants, accessed July 12, 2021.

[8] Information Commissioner's Office and Alan Turing Institute, *Explaining Decisions made with AI,* https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence/, accessed July 12, 2021.

[9] Article 29 Working Party, *Guidelines on transparency under Regulation 2016/679 (WP260),* http://ec.europa.eu/newsroom/just/document.cfm?doc_id=48850, paragraph 27, accessed July 12, 2021.

[10] UK Anonymisation Network, *Anonymisation Decision Making Framework,* https://ukanon.net/framework/, accessed July 12, 2021.

[11] Article 29 Working Party, *Guidelines on Personal data breach notification under Regulation 2016/679 (WP250rev.01)*, 6 February 2018, https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=49827, accessed July 12, 2021.

[12] Information Commissioner's Office, *How do we do a DPIA?* https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/how-do-we-do-a-dpia/, accessed July 12, 2021.

[13] *Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.*

[14] European Data Protection Board, *Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects*, 16 October 2019, https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-22019-processing-personal-data-under-article-61b_en, p. 15, accessed July 12, 2021.

[15] Article 29 Working Party, *Opinion 15/2011 on the definition of consent*, 13 July 2011, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2011/wp187_en.pdf, p. 8, accessed July 12, 2021.
[16] Article 29 Working Party, *Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC*, 9 April 2014, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf, accessed July 12, 2021.
[17] Jisc, *Student Voter Registration,* https://www.jisc.ac.uk/student-voter-registration, accessed July 12, 2021.
[18] Seamless Access, *How Federated Authentication Works*, 8 June 2020, https://www.youtube.com/watch?v=wjvC_PUj4CI, accessed July 12, 2021.
[19] GEANT, eduGAIN Policy Framework, https://technical.edugain.org/documents, accessed July 12, 2021.
[20] UK Access Management Federation for Education and Research, *Rules of Membership*, 12 February 2019, https://www.ukfederation.org.uk/doc/rules-of-membership, accessed July 12, 2021.
[21] REFEDs, *Entity-Categories Home*, 15 March 2021, https://wiki.refeds.org/display/ENT/Entity-Categories+Home accessed July 12, 2021.
[22] S.J. Ahern, The potential and pitfalls of learning analytics as a tool for supporting wellbeing, *Journal of Learning and Teaching in Higher Education* **1**(2) (2018). doi:10.29311/jlthe.v1i2.2812, accessed July 12, 2021.
[23] N. Sclater, *Learning Analytics Explained*. Routledge, London, 2017.
[24] R. Ferri-Garcia and M. del Mar Rueda, Propensity score adjustment using machine learning classification algorithms to control selection bias in online surveys, *PLoS One* **15**(4) (2020), e0231500. doi:10.1371/journal.pone.0231500, pp. 1–2, accessed July 12, 2021.
[25] A.N. Cormack, A data protection framework for learning analytics, *Journal of Learning Analytics* **3**(1) (2016), 91–106. doi:10.18608/jla.2016.31.6, accessed July 12, 2021.
[26] A.F. Wise and D.W. Shaffer, Why theory matters more than ever in the age of big data, *Journal of Learning Analytics* **2**(2) (2015), 5–13. doi:10.18608/jla.2015.22.2, accessed July 12, 2021.
[27] A.N. Cormack, Downstream consent: A better legal framework for big data, *Journal of Information Rights, Policy and Practice* **1**(1) (2016). doi:10.21039/irpandp.v1i1.9, accessed July 12, 2021.
[28] S. Barocas and H. Nissenbaum Big data's end run around anonymity and consent. in: *Privacy, Big Data and the Public Good*, J. Lane et al. (ed.), Cambridge University Press, New York, 2014, pp. 44–75.