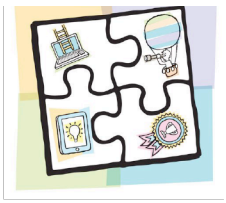


Learning Analytics and Data Protection in Ireland

1. Introduction

This is an introductory consideration of certain issues relating to the implications from and for learning analytics from a data protection (and ethical) perspective.

In particular, it highlights some of the issues that an educational institution would have to consider in greater detail when considering whether or not to adopt a learning analytics programme and or related functionality.

It also considers some of the nuanced data protection issues and concepts.

This document and the content is an early stage consideration and obviously further and or more particular points, nuances and consideration would apply when more details are available when considering a specific proposal or proposed project, such as a more detailed understanding of the existing types of databases, existing data categories and existing data use purposes as apply to the particular institution.

Depending on the particular circumstances, some learning analytics may be permissible, and some not. A careful considered analysis is required in each instance.

2. GDPR

One consideration relates to how the new EU General Data Protection Regulation (Regulation 2016/679)(GDPR) may apply to a learning analytics project in Ireland, and how changes may apply from the previous data protection regime in Ireland, namely under the EU Data Protection Directive 1995.

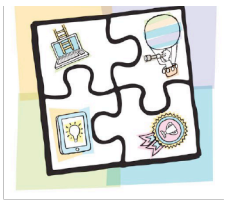
(Specific consideration would also be required of national data protection measures, and as these may be amended on foot of the new GDPR).

3. Data Protection Concepts

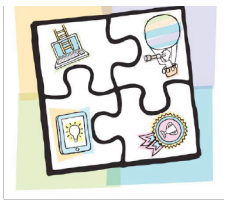
The GDPR sets out a number of important definitions or concepts central to understanding the data protection regime. These definitions are the building blocks of data protection and data protection compliance. These are relevant considerations.

Some examples include the following definitions,

"personal data"	means any information relating to an identified or identifiable natural person ("data subject"); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;
"processing"	means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording,



	organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;
"profiling"	means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements;
"pseudonymisation"	means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person;
"controller"	means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law;
"processor"	means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller;
"recipient"	means a natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not. However, public authorities which may receive personal data in the framework of a particular inquiry in accordance with Union or Member State law shall not be regarded as recipients; the processing of those data by those public authorities shall be in compliance with the applicable data protection rules according to the purposes of the processing;
"third party"	means a natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the direct authority of the controller or



	processor, are authorised to process personal data;
"consent"	of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her;
"personal data breach"	means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed;
"data concerning health"	means personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status;

Please note that there are two main types of personal data, namely (1) general type personal data; and (2) sensitive type personal data. There are more compliance obligations and restrictions in order to be able to collect and use sensitive type personal data.

For present purposes, institutions would need to be aware that health data or data concerning health falls into the category of sensitive type personal data. Hence, the obligations and criteria for data use are more stringent. The present consideration does not analyse whether any of the potential source databases in a given institution or being considered by a given institution contain health data or not. This is, however, clearly flagged as an issue to be considered in greater detail).

A final answer as to whether something may be permitted, or not, will involve up front identification and consideration of health data issues.

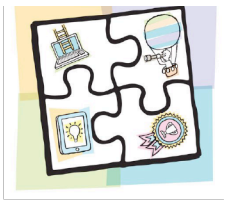
Issues of onward transfer of the data, including cross border transfers are not considered here, but may be relevant in proposed specific instances.

There are also definitions for genetic data and biometric data. Consideration of whether any specific institution's proposal encompass genetic data or biometric data may need further review. More detailed consideration of these specific issues is required.

4. Principles

All personal data collection and processing must comply with what are known as the principles or data protection principles. These are set out in Article 5 of the GDPR. It states that personal data must be,

- processed lawfully, fairly and in a transparent manner in relation to the data subject ("lawfulness, fairness and transparency");
- collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for



archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ("purpose limitation");

- adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ("data minimisation");
- accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ("accuracy");
- kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ("storage limitation");
- processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ("integrity and confidentiality").

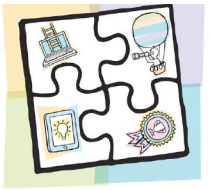
The controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 ("accountability").

5. Lawfulness of Processing

Article 6 of the GDPR refers to lawful processing or the lawful processing conditions. Organisations must be able to comply with at least one of these in order to be able to collect and process personal data – in addition to complying with the principles above.

Processing shall be lawful only if and to the extent that at least one of the following applies,

- the data subject has given consent to the processing of his or her personal data for one or more specific purposes;
- processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;
- processing is necessary for compliance with a legal obligation to which the controller is subject;
- processing is necessary in order to protect the vital interests of the data subject or of another natural person;
- processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;



- processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child. (This point shall not apply to processing carried out by public authorities in the performance of their tasks).

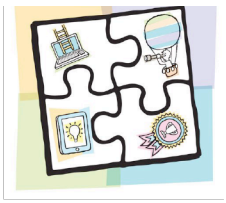
So while consent can be critically important, depending on the circumstances, the legitimising factor may be a factor other than consent. There is official authority even prior to the GDPR that consent should not be overly relied upon if there is some alternative or more appropriate legitimising mechanism. This will be just one consideration for institutions depending on what model and what nuances of learning analytics they may be proposing or considering. It is not a guarantee that consent can be ignored or bypassed. The individual model proposed will be important in identifying which legitimising route needs to be followed. (It may be important to consider also that learning analytics can involve a new or secondary proposed purpose, separate to any primary purpose. Just because purpose 1 may have a legitimising mechanism does not mean that as purpose 2 (e.g. learning analytics) is covered or permitted under the purpose 1 justification. A separate new legitimising mechanism may be required depending on the circumstances).

Member States may maintain or introduce more specific provisions to adapt the application of the rules of the GDPR with regard to certain processing for compliance by determining more precisely specific requirements for the processing and other measures to ensure lawful and fair processing including for other specific processing situations as provided above. We may not know the full details of specific final national measures in addition to the GDPR for some time.

There is also reference to public interest issues. It is yet to be determined whether these would or would not apply to institutions considering learning analytics. Further analysis would be required.

There is also express reference to factors that may need to be considered in specific detail as regards whether a new data processing purpose may or may not be permitted in addition to an initial permitted data processing activity. Institutions will need to identify each of the core data primary databases and systems in question. They will then need to consider what data is collected on each, for what purpose (or purposes) and what legitimising mechanism is involved. One of the issues will be whether leaning analytics are closely linked to and transparent to students as a follow on from the initial data collections and data processing purposes. The further away the learning analytics activity is from existing activities, the more it needs to be expressly made known to students in a transparent manner, and the greater the potential need for consent and other options made available to students. Data protection considerations (and indeed ethical considerations) arise where a new activity is so unrelated to an original activity that a reasonable student could not reasonably anticipate the new activity of processing arising in relation to their personal data or as a closely linked follow on activity.

It may be that an institution concludes that learning analytics, or at least certain learning analytics, is so new or so different from prior activities that it cannot be included in a



prior consent or a prior legitimising mechanism.

In those instances, it is necessary for institutions to consider how a new consent or alternative legitimising mechanism can apply to new learning analytics activities. It should not be automatically considered in such an exercise that a generic label of "learning analytics" is a sufficiently defined and understood concept that students could consent to same, nor that seeking to provide necessary transparency would be adequate with such a general term. More details and information may be needed.

An institution would need to be very specific in identifying each individual purpose being proposed under a given learning analytics project. All purposes must be individually considered in a diligence exercise. All (permitted) purposes must be transparent. This will require such purposes to be individually documented and individually considered. It may be the case that a new proposed data processing purpose 2 may be permitted but not new proposed data processing purpose 3.

6. Individual Consideration is Needed

The individual considerations may differ from project to project and from institution to institution. Very particular diligence is needed by each individual institution as regards the potential learning analytics projects being considered.

The initial proposed activities for institution 1 may be A and B, whereas institution 2 may be considering purposes A, D and E. Therefore, the considerations and diligence exercise of each institution will be different.

This also highlights the need for each institution to carefully document and outline the proposed activities and purposes of the proposed individual learning analytics project in advance. It is also necessary to identify and document the purposes, technologies, configurations and databases being proposed.

Just because institution 1's proposed project may be permitted after diligence, does not mean that the project being proposed by institution 2 is permitted after diligence.

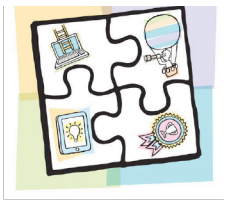
It is important to realise also that just because a particular learning analytics system may be available or in use elsewhere does not mean that the same system is permitted in Ireland (or permitted the EU). Some jurisdictions may have different, even lesser, personal data and data protection standards and safeguards than the EU.

7. Modelling

One issue highlighted relates to the use of learning analytics capabilities but strictly limited to (data from) databases and systems in such a manner and configuration that NO personal data is involved and NO individual students can be identified.

It has previously been understood that so long as given research activities (in this instance modelling) does not contain any personal data at all, it falls outside the data protection regime. This is still the position under the new GDPR.

It has also been considered that a given data set which contains personal data, but which has all personal data extracted and filtered out, known as anonymised, that the new filtered data set is cleaned and possible to process for research or possible other activities.



However, there are potential practical difficulties. There are an increasing number of examples of where organisations thought that they had successfully filtered and anonymised a data set, but other researchers (or reporters) have been able to find personal signals still remaining in the dataset which do lead to identification. The entire data set is therefore potentially still personal data, and absent lawful legitimisation, the dataset cannot be used as was intended. The exercise has been unsuccessful.

Given that there are ever increasing potential signals or identifiers in a whole host of different databases, it is increasingly difficult and involved to successfully anonymise the intended research or modelling database. Researchers now need to consider each input database on a case by case basis, as well as all immediately identifying characteristics or signals, and now also any potential signals or identifiers that would not jump out in the normal course presently, but which could be used in future by third party researchers, students, students representatives, or officials to identify students. It is important to bear in mind that personal data can be identifiable data, or data which becomes identifiable once combined or linked with other data.

The emphasis on anonymisation issues is more expressed under the new GDPR than previously.

Potentially at a headline level, so long as there is no personal data involved and no individual student can be identified by the system, a proposed learning analytics model, such anonymised modelling, may be permissible under data protection grounds. However, the specific details would need to be considered and no assumption should be made in advance.

In terms of nuance, the differentiation between where the data collection and processing system is linked directly to the live student source systems and databases (A – B), from a system where the relevant data is copied, anonymised and filtered before being transferred into a new clean database (A – B – C) may need to be considered.

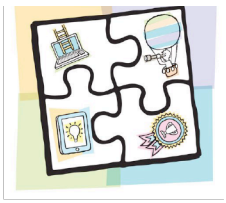
The ethical and risk issues which arise are that A – B – C is safer than A – B. There are more risks with A – B and despite good intentions, an unanticipated leakage can arise where personal data or a personal data identifying signals are carried into the modelling stem. Thus, the modelling system is contaminated. These risk issues are possible through system design, bad design, haste, unanticipated carry over, unanticipated signals, human error, lack of training, etcetera.

Educational institutions might like to consider whether A – B – C is more preferable as being more permissible, ethical and more risk adverse.

In terms of data protection, while both avenues might be permissible, after appropriate diligence, there is greater risk that if just one issue goes wrong, the entire system falls into breach of the data protection regime. There are also other consequences, such as data breach, official investigations, complaints, etcetera.

8. Live Learning Analytics for Identified Students

One issue raised relates to the consideration, if not proposal, by an institution to use live learning analytics capabilities in relation to live databases and systems which include the personal data of real identifiable students.



In this instance there will clearly be certain personal data. It is critical to beginning a query such as this to clearly document and then consider the following,

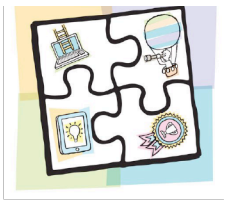
- Proposed purpose; or
- Proposed purposes, and in such case each purpose individually;
- What type of system is envisaged;
- What data exists and in what databases;
- General personal data and or sensitive personal data;
- Consent;
- Other legitimising ground;
- Transparency;
- How linked and connected each proposed purpose is compared to each original purpose and each original database;
- Controlled data access;
- Labelling;
- Student rights;
- Data deletion and life cycle.

The issue of labelling could be quite important on a number of levels. It may not be the wisest decision to label some of the identified students under a learning analytics system as "likely to fail" or similar. There are obvious ethical issues. In terms of data protection such a label tends to escalate the personal sensitivity type of the label to the individual student. If contacted with such a label, such a label could have personal esteem issues, labelling issues, emphasise and re-enforcement in a negative way, etcetera. The risk of data leakage are emphasised also with such a label. There are also issues in terms of such a label requiring potentially stricter access controls. Students' right bodies (and others) may be less enthused with such labelling when informed of the existence of it. This may raise policy issues in and of itself.

9. Access to Data

It is important to consider, all other issues resulting in a possible diligence exercise, the issue of data access and access control. It is required that access to this type of personal data (i.e. predictions in relation to identified students failing courses) be documented in a policy, limited and restricted. There should be defined hierarchy of access to the report prediction data. It may be that not every lecturer is required to have access. It may be that a head of department or a designated "student learning" officer in each department only, would have access. In this scenario, there would be defined documented procedures as to what they may do with such data, whether they are permitted to disclose it to head of department, individual lecturers, the individual student, etcetera.

If there is a case where alternatively lecturers were per se permitted access, it might be that procedurally, technically and in training, it is clear that teaching assistants or



demonstrators or similar more junior persons are not permitted access. Technical, policy, procedural and practical safeguards must be applied.

In terms of access generally, it is very important that not just the obvious aspects of each source database and the amalgamation of relevant parts of these databases for the learning analytics functionality, are considered for data leakage or unauthorised access, but also that a detailed exercise is undertaken to anticipate and eliminate non-obvious leakage. It should always be borne in mind that even where all other aspects of a project may be data protection compliant, if just one aspect is non-compliant, or becomes non-complaint, that potentially undermined the entire project.

If access and access controls are too loose, or worse not in place (or slightly less worse, documented but not enforced whether in practice or technically), it may be safe to assume the entire project falls foul of data protection (and ethical) rules. Such a project should not proceed, or if already initiated may have to be immediately discontinued.

10. Student Rights

Organisations also need to be aware of student rights, as they exist for all individual data subjects. One of the data rights is the right to obtain a copy of their personal data. This is known as the data access right. The extent to which their personal data and records exist in a learning analytics system, such date may have to be produced when a student requests same. They may request such data specifically, or it may fall to be included in a more generally request for their personal data on all files and systems. This is referred to as a data access request.

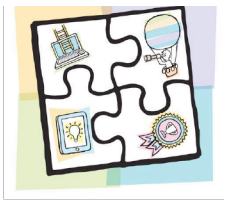
Notwithstanding consent and other legitimising grounds, there are also other rights issues to consider. Individual data subjects have the right to obtain from the controller the restriction of processing in particular circumstances. This will need to be considered when these projects are being proposed in greater detail.

Individuals also have a right to object to automated individual decision-making in particular situations. Article 21 of the GDPR states that “The data subject shall have the right to object, on grounds relating to his or her particular situation, at any time to processing of personal data concerning him or her which is based on point (e) or (f) of Article 6(1), including profiling based on those provisions. The controller shall no longer process the personal data unless the controller demonstrates compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject or for the establishment, exercise or defence of legal claims.”

It may be that there is merit in proposing a system which operates in an automated manner, with as little human or operator input as possible in terms of day to day running of the system. Some AI systems seek to operate on this basis. However, there is sometimes controversy in relation some example of proposed AI or automated systems.

The analogy is that depending on how a proposed learning analytics system is configured, it may be considered to be intrusive or too personally intrusive. On occasion, some of these issues might be reduced in severity by an automated system. The appropriate balance would need to be carefully considered in each instance in advance.

This can also trigger other data protection issues. For example, Article 22 of the GDPR



refers to automated decision (including profiling). The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her. A typical example often used is an adverse credit decision being arrived at in an automated manner without human input. One of the rationales is that a very important decision may be made adversely affecting the individual but there may be an error in the system which a human supervisor would have prevented. Institutions might have to consider the type of labelling system being proposed, how widely the system was accessed, how interventions may occur and by whom, transparency, etcetera, and ultimately the affect on the individual. Are students identified? forced to take extra classes? What are the classes called? Where are they? Do other non-triggered students know of these additional classes? What is an "effect" and what is a relevant effect in the context of learning analytics systems?

Frequently students seek references from lecturers. Might a reference be coloured by a history of an individual student having at one stage being triggered via a learning analytics system? There are data protection (as well as potential ethical) issues to consider in greater detail. These diligence issues need to be considered. Where permissible, appropriate related policies, documentation and training need to be developed.

In addition to considering the most appropriate legitimising mechanism, and associated transparency issues, institutions may also need to ensure layered transparency. Notices in relation to the system may have to appear in a number of different locations, and possibly in a number of different text lengths and detail. Opt-outs may also have to be considered.

It may also be the case that different option mechanisms need to be presented, and remain available, to students.

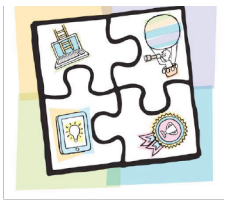
Even where a particular learning analytics system and particular functionality may be decided upon, after diligence and otherwise being permissible, it may be appropriate that the system is only applied to year 1 students, and on a rolling annual basis, thus excluding students already in the system.

Depending on the diligence outcomes, one would separately consider the issue of new students and existing students. This may include registration and re-registration models as appropriate. One of the considerations may be legitimisation, transparency, intrusiveness, and how linked or connected the learning analytics system is to pre-existing data processing purposes.

11. Changes

Where the purposes and activities change, the institution may need to consider whether the potential changes, addition, expansions, etcetera, are permitted. Just because purpose 1, or purposes 1, 2 and 3 may have been permitted after earlier diligence, it may not be the case that new purposes 4, 5 and 6 are automatically permissible. Individual consideration of each additional purpose is required.

12. Security



It goes without saying that security and data leakage prevention are important. It is difficult to overestimate the importance. Educational institutions may be expected to apply greater diligence, data protection and ethical standards than at least some other organisations and commercial companies. They should be expected to lean more towards data security, best practice, risk adverse, etcetera, than more risk friendly companies. Part of the diligence exercise may mean recognising appropriately that more than one system and more than one pre-existing student database is involved. The security measures for learning analytics may need to surpass existing security levels for an individual database or system. The nature of the content, and prediction functionality may separately emphasise the need to surpass existing normal security.

13. Retention

Where a learning analytics system is permitted and decided upon, the institution needs to consider how long it will store the data. Personal data should be kept no longer than is necessary. So data should have a natural life cycle and be deleted after a reasonable period after the purpose has been achieved. This might be after the student has left university, successfully graduated or possibly even sooner, e.g. after a particular course is completed. There should be an appropriate policy identifying the issues, consideration and setting out the policy in terms of deletion. This is additionally important given the sensitivity that may attach to the learning analytics data for certain students.

14. Monitoring

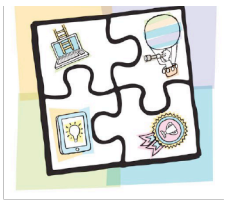
Issues of monitoring are generally frowned upon in terms of data protection and are not permitted, or at least only under exceptional circumstances. This should be considered generally by an institution in terms of a learning analytics system. However, it needs added consideration where it might be proposed that the email system, headers, metadata (or content of communications) might be incorporated into the learning analytics system. Such use could be taken to lean towards the system amounting to or using monitoring of students in their everyday university electronic activity. The more all embracing, intrusive and detailed the sources and inputs to the learning analytics system, particularly email, may indicate problematic issues from a data protection point of view.

15. Conclusion

While learning analytics systems may potentially be permissible in Ireland, there should not be a rush to roll out such a system without appropriate data protection and personal data diligence. There is not automatic presumption that any such system if automatically permissible or permitted. As per best practice, data protection should be incorporated on day one of such proposals and considerations, as opposed to being raised at the end of the project or worse when a go-live date has already been set. This would increase a risk trajectory of something going wrong and falling foul of data protection or ethical best practice or permissibility.

In terms of progressing a more particular and informed discussion of the data protection issues applicable to a learning analytics system in Ireland, an institution should,

- identify a list of the typical databases in an university which contain student data;
- identify which of these might potentially be used for a learning analytics project;



- identify all of the individual signals or categories of student personal data held on each such database;
- identify all of the potential individual purposes that a learning analytics system in Ireland might be used for.

The above should not be taken to be recommending any particular system, nor that any such system is permissible or recommended. Obviously individual institutional diligence and individual considerations arise in every instance.

Dr Paul Lambert

lex@mydistillex.com
Author:
Data Protection Law in Ireland (Clarus Press);
A User's Guide to Data Protection (Bloomsbury);

The Data Protection Officer, Profession, Rules and Role (Routledge, Taylor and Francis)