OfLA Project
2018-1-UK01-KA203-048090

O15 – Guidance on using Institutional Data

RESPONSIBLE PARTNER:
NOTTINGHAM TRENT UNIVERSITY

PARTICIPATING PARTNERS:
UMC UTRECHT
ARTEVELDEHOGESCHOOL

Output 15 – Guidance on using institutional data

This report will complement O14. It is intended to provide some background context and support the processes identified in O14. It is likely that this report will focus more on the IT infrastructure and processes needed to support effective tutor interventions.

The work will be carried out by the project team across all three years. We will particularly use the middle transnational meetings to review institutional data challenges and use these as the basis for the work.

A1. As part of interviews, observations, and analysis of outcomes as we carry out O6, O9 and O12 and from the questions we ask in O5, O8 and O11.

A2. At our fifth transnational meeting, we will review the data and edit the document.

A3. We will test these resources with staff involved, and at external events and upload them to the project website.

"The European Commission support for the production of this publication does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein."

This output is a result of the European Erasmus+ project OfLA (2018-1-UK01-KA203-048090)

Co-funded by the Erasmus+ Programme

Table of Contents

Introduction and background

This output provides an overview for colleagues of the institutional data issues associated with the use of learning and analytics and early warnings. It is derived from the operation of learning analytics alerts and survey work at each partner. Advice and guidance will not be specific or overly technical; this is intentional as so it can be applied to a variety of institutions, and in a variety of different contexts and processes. In this document, examples are given relating to various systems currently in use at each institution, and therefore some contextual information about the data infrastructure is helpful. This output is specifically designed for a range of staff members who work with data at an institution including those in institutional IT departments, researchers and senior managers.

How data is used at the partner institutions during the project

At Arteveldehogeschool, no specific learning analytics platform is in place. Student data is used however as part of the student tracking system (SVS), which includes both objective data about the student (such as background, attendance, grades), as well as feedback produced by the student (such as the FIT-test).

At Nottingham Trent University, a learning analytics platform, the Student Dashboard, collates data relating to a student that is used as proxies for 'engagement'. Daily engagement ratings on a five-point scale are produced on a daily basis for students, presented on the Student Dashboard, that is accessible by both staff and students, used primarily in the support process.

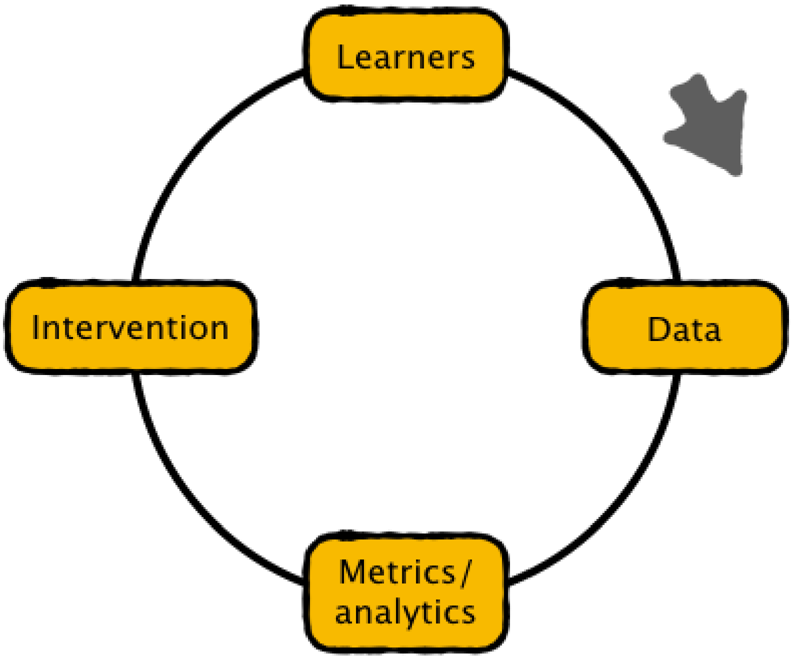
At University Medical Centre Utrecht, personal data and progress data is stored and presented on the learning management system (Osiris). Some data regarding student progress, engagement, and behaviour is gathered, however this is a manual process conducted by individual staff members. As a result, this institution is currently in the process of developing their own learning analytics platform.

For more context about the various systems and policies in place at each of these institutions please refer to the O6 case studies for each institution, provided on the OFLA website.

It is also important to consider what data is being described within this output and within the research project more widely. Data systems at each OFLA institution, although very different in nature, follow a similar set of rules in which they fit in with the supportive process. This is best represented using the Clow (2012)[1] cycle. At all three institutions, learners produce data, which is captured, analysed, and understood in some way, before being used as part of the support process. As a result, the student should (if successful) change their behaviour, and this change in behaviour can be illustrated in the student data. Subsequent analysis of this data should therefore identify whether the intervention is successful, or further intervention is necessary. This cyclical process as described by Clow is illustrated in figure 1 below.

[1] Clow, D. (2012, April). The learning analytics cycle: closing the loop effectively. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 134-138).

Figure 1: The Clow Cycle



The way in which staff specifically use data in the support process is detailed in the OFLA three stage model of support, referenced throughout this project. This details how the steps in between the data and the change in behaviour; data (in some form) creates an 'alert' or 'trigger for action', the staff member then finds a method and approach for communicating this to the student, and finally an intervention or some action takes place in order to change student behaviour. This model is detailed in figure 2 below.

Figure 2: The OFLA three stage model of support

Methodology

The guidance and recommendations detailed within this report has been produced firstly by reviewing what was learned throughout all previous case studies as part of this project. A literature review conducted at the start of the project (described in O4) provided a foundation for subsequent research as part of this project. In the first year of this project, the researchers investigated the current supportive process at their respective institutions (O6 case studies). Subsequently, the researchers conducted pilot case studies in the second year of the study (O9 case studies) and wider larger scale research pieces in the third year of the project (O12 case studies). Findings from all of these outputs provide a rich and detailed picture of the issues that data handlers may experience, and potential solutions that may wish to be explored. Throughout this guidance, these findings, issues, and potential solutions are referenced in brackets detailing the institution, output number (with specific case study), and section or page details.

In addition to reviewing what was learned throughout the research projects interviews were conducted with 4 members of staff from the 3 OFLA institutions. These were informal interviews and focused on both their experience of working with institutional data during this project, as well as prior to the commencement of this research project. General questions were asked, such as where they or their institution first started using data as part of the supportive process, how this data was selected and presented, what they would like to use if they could, examples of data problems they have experienced, and what their own recommendations are based on their detailed experience. These interviews lasted approximately 1 hour, conducted through Microsoft Teams, and were not recorded and therefore no direct quotes are used in this output. Where guidance is based on these interviews however, the example will note that it is based on interviews conducted for this output specifically. Each of these are therefore considered ‘micro-case studies’ relating to each institution.

Finally, throughout this output, examples provided relate to each OFLA institution. The names of each institution will be abbreviated for ease; Arteveldehogeschool (AHS), Nottingham Trent University (NTU), and University Medical Centre Utrecht (UMCU).

Guidance and Recommendations

This project primarily focuses on the use of data to support students. However, research also provided an insight into what may need to be considered when embedding the IT infrastructure behind the data.

Collating this learning into a single piece, this document provides guidance to an institution wishing to develop data and/ or learning analytics systems as part of the student support process. It is designed for staff members who work with data, conduct research, or consider systems at a senior level. Advice and guidance is not detailed or overly technical; this is so it can be applied to a variety of institutions and differing contexts.

Recommendations are grouped into overarching themes. These themes are not necessarily chronological in nature, instead highlight different areas to be considered before developing a system. Each of these themes are described below.

1. **Purpose and Outcome Focus:** The long-term impact of a data system, how its outputs are used and who this system will designed for, ultimately dictates the structure of the system itself.
2. **Type and Availability of Data:** Learning analytics/ early warning systems will be profoundly shaped by the quantity, quality, timeliness, and ethical dimensions of the data available.
3. **Processing and Presenting Data:** How data is treated and displayed to the user can be achieved in a much more efficient and effective way if the IT infrastructure itself compliments the process.
4. **Data Sourcing and Storage:** In addition to identifying what data to use and understanding how to process it, users must understand where the data comes from and where it is held before and after it is processed.
5. **Timing and the Academic Cycle:** In addition to what data is used, understanding when the data can and should be acted upon, determines the structure of the system and process.
6. **Stakeholders and Data Users:** Throughout the project, staff members from across the university and beyond were consulted as both providers and users of data, and their contribution was found to be critical.
7. **Ongoing Development and Refinement:** The majority of this guidance applies to the initial development of IT infrastructure and the early stages of managing institutional data; however, there are continuing factors that an institution must consider on an ongoing basis.

1. Purpose and Outcome Focus

Several questions should be asked during the initial planning stages of using 'big data' for student support. Most importantly, establishing outcomes, the method in which these outcomes will be achieved by the user, the extent to which the data informs these decisions or provides solutions, and the extent to which data literacy is needed by the user throughout this process.

Breaking this down in more detail, we **recommend that you consider:**

1.1 The outcomes of using data in support

The learning analytics platforms we use have a clear and statistically proven predictive power. We know that the indicators we use (whether that is proxies for 'engagement', belonging, participation in other aspects of university life, attendance, performance, or an individual's own self-reported views), can all help determine the likelihood that a student is going to fail or withdraw during their time at university.

Simply analysing this data after the fact can provide an insight into relationship between behaviour and performance of a student, however any way to use that data to help that particular student is limited and occurs only after the student has already failed or withdrawn. Learning analytics allows us to understand student behaviour and performance while the student is still enrolled and has yet to complete their course. Therefore, action can still be taken to change or improve performance, due to the predictive nature of the behaviour. At all three OFLA institutions, responsibility for improving performance lies with the student, however those staff members key in supporting this process differ between institution. There are also differences in the level to which the student is treated as responsible or needing to be guided with support, and to the extent that student withdrawal is seen as entirely negative problem. Who is involved in the supportive process, when this occurs, how this occurs, and most crucially, what the outcomes of support should look like, determine what data is needed for the learning analytics system that you wish to create and embed. Different sets of data are needed if you are looking to change behaviour, or change performance, or change study technique, or change attitude of the student. Considering the outcome first, allows you to better understand what data you will ultimately need to use.

At AHS, a specific learning analytics platform has not been designed. Instead, data is included in the SVS (student tracking system), with the aim of providing supporting information as part of a complete overview of the student (AHS O6, page 2). At NTU, a specific learning analytics platform has been designed, and therefore this data is central to the system and its function (NTU O6, section 2.1). In one system, data **supports** the overview, and therefore specific and key bits of information are the most effective pieces of information to present. In the other system, the data **provides** the overview, and therefore a significant amount of data that paints as complete a picture of that student as possible, is needed. This point is referenced from a policy perspective in our output describing policy recommendations (O14, recommendation 3).

1.2 Why the user may need to use this system

In order to achieve the outcomes as defined in the initial planning stages, one must ask what user needs to access the data. If data is to be accessed by staff members or the individual student only, would this change the type of data presented, in comparison to a system accessed by both staff and students. At all three institutions, the data and/or learning analytics platform is accessed by a mixture of students, tutors or study advisors, course delivery staff, policy makers, and research staff. All three institutions believe that a) the responsibility for change is with both the staff and student, and b) that an open and transparent approach to data is required for a better relationship between student and institution (O6 case studies).

There are however some differences between OFLA institutions in the users of the data, that have led to differences in the data that has been used in the system. For example, at UMCU, an developmental direction for a new learning analytics platform would be for a research project coordinator to view student data in order to enhance feedback provided by rubrics (UMCU O6, 2.4.3.1). In this example, detailed data of the feedback received by the student, including when, how, and what was delivered, is needed. This allows the staff member to provide support in **what** students may need to engage with going forward. At NTU, staff use data to understand and support students in **how frequently** they are engaging in studying, and the method of that engagement, rather than reflecting on what was produced as a result (NTU O6, 2.1). Data here is more constructive therefore in identifying students who have stopped engaging in study, as opposed to using data to identify how engaged students can perform better.

Ultimately, these users crucially need to be able to act on what that platform shows them, and therefore data needs to be relevant to that user.

1.3 The extent to which decision-making is determined by the system

Through our research and through discussions with researchers at each of our institutions, there is a consistent agreement that students have a responsibility for changing behaviour in order to achieve student success, remains with the student. Staff have a responsibility for coaching and providing support for that student, and can provide advice and guidance throughout the supportive process. The decision making for what that change might be however can vary; in one extreme, the user is simply instructed on what they need to do by the system, and in the other, are given all the information by the system for them to make their own autonomous decisions. This applies to both students in the actions they take, and the staff members in when and how they are to intervene.

Considering staff interventions, reflective diaries produced by staff in our O9 case study, suggests that staff themselves feel their responsibility is to interpret data and for them to be able to use their own expertise and rapport with the student in order to provide guidance (NTU O9 'staff reflective diaries', sections 4.1 and 4.2). Some guidance on what constitutes high vs low engagement however helps the staff member understand when action is to be taken, and at times, the data itself highlights students who require further contact (NTU O9 'Mid-term Reviews'). The Dashboard system also makes some decisions for the staff member in contacting a student, and research has been conducted through this project as to specific timeframes for objective data to require action by the staff member. In the latter case, the decision to act is made largely not by the staff member, but by the system itself (NTU O9, 'reducing the alert timeframe'). Similarly, at AHS, different data sets are used to help staff members make a decision on what action should be taken in the form of a FIT test, or whether the data directs the staff member to take action, in the form of the BEM programme attendance (AHS O9, section 2.3). Subsequent research by AHS investigates further the autonomy and self-direction of the students (AHS, O12), and again, the concept of where decision making is needed for the student, in conjunction with the data they receive about themselves, features heavily throughout the OFLA project as a whole.

The extent to which decision making is made appear to be on a sliding scale from instructing actions to complete autonomy, with data either directing action or helping to make an informed decision. This concept is similar to hierarchical models of knowledge or learning. An example is 'Bloom's Taxonomy', describing various levels of complexity and specificity of learning. Applying this to a learning analytics platform, a system that has low autonomy for the user, has the advantage of simplicity in the information it presents. The data needed therefore can be simple in its collection and/ or presentation (for example, 10% attendance, suggesting "this is bad", and therefore the student must do x number of hours extra work). Conversely, a system that has high autonomy for the user has the advantage of the user being able to add context and expertise to a unique situation. The data needed therefore can be complex, detailed, giving the user a vast amount of information (for example, detailed notes of how the student studies, suggesting "this is potentially bad, but there are also reasons behind this, and therefore the staff must guide the student with a bespoke plan of action").

The latter example of greater autonomous decision making and therefore more complex data has some issues, for example data literacy, which we will discuss next.

1.4 Improving data literacy of users and stakeholders

A learning analytics platform that provides more autonomy for the user, is advantageous in that it can use the tacit knowledge and contextual information that the individual has, that the system would not be able to consider. To differing extent, all users of data systems at the three OFLA institutions, whether they are students or staff, have to make decisions on actions taken as a result of the data they see. Whether it is feedback from an assignment, grading or results, attendance, motivation, academic readiness, or engagement, staff and students need to be able to identify **when** action should be taken, and **what** action should be taken.

Elements of systems at the three institutions contain more complicated information, that does not instruct, but rather, informs. Students at NTU see their own engagement on the Student Dashboard, and are encouraged to change their study behaviour sometimes independently of staff members (NTU O6, section 2.1.3). Similarly, students at UMCU can see their own progress through the exist system, and support is sometimes reliant on the student putting themselves forward for support as a result (UMCU O6, section 2.1). In both of these examples, the student hasn't just understood that the data indicates they may be at risk of failure, but that they need to take specific actions to correct this.

Systems with more complex learning analytics data require a higher level of understanding of the data that it contains; as the complexity increases, the more the user needs to know **why** the data is important, not simply whether the data represents a positive or negative outcome. This therefore requires data literacy to become an increasingly important component to a learning analytics platform. From providing descriptions along with the data (NTU O9, 'Mid-term Reviews', section 2.2), to 'onboarding', to training, to contextualising the information more widely, data literacy is fundamental to the user. Again, in addition to the impact this issue has on the data, policy is also required to ensure the data is understood at all levels (O14, recommendation 4).

2. Type and Availability of Data

Although we may have idealistic outcomes when planning to use data, an institution has to realistically consider what data it can and should use as part of a learning analytics platform or student support system. This does not necessarily mean that an institution has to rely solely on existing data points, but it does need to consider what these data points are, how many should be used, and how many are to be included in the support process overall. In breaking this down further, **we recommend that you consider:**

2.1 The specific use of dynamic data over static data

For users to be able to identify problems that lead to a student being at risk of failure, understand the causes of these issues, and make changes to ensure the student or individual does not remain 'at risk', the data needs to be 'actionable'. Demographic data or entry grades, for example, can be a good predictor of subsequent academic performance at University. An example of this found during the OFLA was in the regression analysis conducted by UMCU on student background data at the start of masters education (UMCU O9, section 4.2.1), finding a clear predictive link between previous academic achievement or even the age of the student, and the likelihood of failure or withdrawal. Similarly, students who are accepted onto a course at AHS from a vocational background in secondary education were again more likely to be perceived as at risk (AHS O6, section 4.2). In both of these examples however, the student cannot change who they are or past performance before they started University, and so we believe there is a practical and moral reason for an institution to not use this type of data in learning analytics platforms.

A continued source of data such as attendance, performance, engagement, behaviour, or student feedback, can change once an intervention has taken place. This would be considered a dynamic data source, and one that would be effective if the learning analytics platform is used as part of our three-stage model of support. The student can be informed during the comments that not only is dynamic data much more useful in terms of identifying **when** to act during the 'trigger' stage (NTU O6, section 3.2.1; NTU O9 Reducing the alert timeframe) but provides evidence as part of the communication stage (NTU O6, section 3.3.1; AHS O9, section 4.1 and 5.3) and subsequently helps to direct and monitor actions as a result, such as targets in attendance and whether this target was achieved by the student (AHS O9 4.2). Dynamic data can be included in action planning, and therefore staff can monitor whether the student is still at risk based on the data (AHS O12, section 3.3.1), and therefore potentially closing this loop.

2.2 How the data points are chosen and reviewed before use

Most higher education institutions collect a massive amount of information about students at various points throughout their student life. Learning analytics uses big data on students to predict outcomes based on previous performance, however considering what data is to be used in your system however cannot be based on predictive power alone. This is because the staff member(s) responsible for choosing what data to use as part of the support process, need to understand the context of **why** that piece of data is important and whether it is appropriate as a predictive tool.

Machine-learning without contextual interpretation can lead to a learning analytics platform that is highly predictive, but with data points that are not able to be changed. An example of this was found in NTU as part of the interviews for this specific output, during the discussion about when the data was initially selected to be part of the Student Dashboard. Analysis of student data with a correlation of subsequent success found one data point that was an almost perfect prediction of student progression into the second year for students at NTU. That data point was however 'enrolment status'; the machine-learning process had identified that the most accurate prediction of whether the student was going to progress to second year, was the process of enrolling on to the second year that takes place prior to the start of term. Human interpretation was required to understand that creating an 'alert' based on a student not enrolling to second year would mean any kind of support would come too late into this process.

Using a machine-learning approach to find data sources that have the most accurate predictive power, and then reflecting on the reasoning and context of that data after the fact, was a bottom-up approach taken by NTU prior to the start of the OFLA project. UMCU have taken a top-down approach when developing their learning analytics platform, by forming hypotheses of what data they expect successful students to produce, and then analysing subsequent performance to understand whether this hypothesis has been met. An example of this was considering performance data throughout a term, and the expectation that they see a positive deviation for this metric. The hypothesis is formed therefore that if there is no change for this metric, or more concerningly, a negative deviation, that the student may be at risk (UMCU O9, section 4.1). This hypothesis was proved true, however the process by which the data was analysed in this way, came about only first through an understanding of what that data point represents.

2.3 The amount of data points used

At all of three OFLA institutions, data is collected about students in various forms and at a wide variety of 'touch points' throughout their University life. Many of these data points could indicate future success, so if one has a variety of data points, why not use all of them? There are several reasons why you may need to be considered and measured approach to the inclusion of data sources.

Firstly, the more data points you use, the more complex the system (the issues with maintaining a complex system are discussed later). At NTU, we have developed an increasingly complex learning analytics system over the past few years, with more engagement proxies being added, as well as different ways in which users can engage with the system. This development was not only due to the subsequent availability of accurate predictive data, but because this is part of a wider change management process. A learning analytics system needs embedding at various levels throughout the institution, and a more complex system can lead to confusion amongst users (NTU O6, section 3.5.3).

Secondly, the more data points used, the more likely a student will register some form of 'positive performance' or 'engagement'. Learning analytics platform that include the use of engagement or attendance data for example, give students multiple ways in which they can register data that suggests they are on track. As these systems rely on an alerts based on a lack of positive interaction, allowing for students to register 'engagement' or attendance in more ways, means there is a greater chance that the student who is struggling may still register a positive interaction, and is therefore hidden from the alert. This issue was considered during the OFLA project specifically, as part of UMCU's work on the development of their own learning analytics platform (UMCU O9 section 4.1.1).

Thirdly, the more data points that are used, means a higher level of data literacy is required and a greater understanding of the information in the system is needed. This links back to section 1.4 earlier in this piece.

3. Processing and Presenting Data

A fundamental part of using data is not simply collating data, but processing and subsequently presenting it on the platform. In this section, we will discuss how data can be presented in different ways and in different circumstances, and how this can affect how it is collected and processed.

Breaking this down further, **we recommend that you consider:**

3.1 How you will present the data

To explore this issue further, we can give the example of attendance, being a metric for success. There are, however, multiple ways however in which you can present attendance. At NTU, the Student Dashboard system lists 'engagement', and one of the metrics for engagement is attendance. Therefore, if the student attendance, this registers a positive interaction for the student, which can be thought of as a data point of "1" for attendance. In this way, the data processing of "attendance" is kept as simple as possible; if a "1" appears against a day, that means the student has attended a scheduled teaching session, whereas if the student does not attend, there is no "1" processed by the system. Although this system is simple, what it cannot easily tell you, is whether the student was absent from a scheduled teaching session, or whether no scheduled teaching session took place.

The concept that an institution may be interested in when a student **does not** attend, rather than measuring when a student **does** attend was explored during the OFLA Project. AHS for example found that absence at intermediate test moments can function as early warnings that a student is at risk (AHS O9, section 6). In order register an absence rather than attendance, the institution must record a data point (for example, a "0" for the day), rather than simply having recorded nothing at all. How the data is presented therefore has a direct impact on the way it is captured.

3.2 How presenting more data can be demotivating

As using big data becomes more embedded within an institution, there is a continuing focus on refining the accuracy of the data and increasing the amount of information available as a result. There are however examples where providing more information, or more accurate information, can have the opposite effect. At AHS for example, it was found that providing students with more data and resources did not always result in students taking action (AHS O9, section 6). Similarly, at NTU, increasing the number of data sources, resulted in some staff becoming more disengaged and frustrated with the system (NTU O6, section 3.5.3).

The issue of presenting more complex but more accurate data resulting in declining motivation was discussed during the interviews specific to this output. An example was given of the NTU Student Dashboard system presenting 'cumulative' engagement scores to students. A cumulative 'engagement' score is a good way to illustrate whether a student who 'fell behind' due to a period of low engagement managed to 'catch up' by engaging at a much higher level. This data is more accurate in presenting the total engagement for that student over a period of time, as opposed to simply taking an average score. The reality for many students, however, was that seeing a cumulative score overall could lead to a student working much harder than their peers, however still being told that they are at risk of failure. At NTU, feedback directly from the student via the Student Transition Survey suggests that motivation is often not objective. This instead relies on 'positive' emotion and a feeling that they can make a difference going forward, rather than through a pressure to catch up, despite this need being true as understood purely in the data (NTU O6, section 4.4). One must consider therefore whether the data being presented is 'motivating' as well as accurate.

3.3 How you access data during the intervention

Through our research within the OFLA project, we understand that data used in learning analytics systems is critical for identifying students at risk of failure, and therefore acting as a trigger. Our research has also suggested that there are positive outcomes to the support process if this data is presented to the student during the communication stage (AHS O9, section 3.2 and 3.3). In addition to this, we understand that it is critical for data to be presented during the intervention stage; be it as an 'external arbiter' during the support session (NTU O6, section 3.3.1), in order to discuss and set targets for the student (UMCU O6, section 2.3), or even as part of the system recording the intervention itself (AHS O6, section 4.4).

There are practical challenges that staff need to overcome therefore in order to effectively use and access data during the intervention. Students who were not able to discuss or process their own data as part of the intervention are less likely to access support (AHS O12, section 4.3.1) and so the data needs to be available to both staff and students at all times. We also understand that interventions in the form of support sessions can take place in a variety of locations on and off campus (NTU O9 Staff Reflective Diaries, section 4.6), emphasising the need for a platform that can be accessed remotely. Finally, the relevant data needs to be accessible without the need for specific software, complex methods of producing reports, or waiting for data to be requested; any of these issues would have made it impossible for staff to make successful data-led interventions as part of a call campaign (NTU O12, section 3.4).

3.4 How to present data with appropriate context

In order to fully comprehend performance data for an individual, both the staff and student need to understand how it fits into the wider cohort, or against their peers within a subject or course. This is part of the wider process of understanding students at risk, as well as the need for improving data literacy for the user. For example, at NTU, each school (or faculty) have different percentages of attendance that act as a trigger for action; 80% attendance may be acceptable for a student in one course, whereas be a cause for concern in another (NTU O6, section 2.2.3).

At NTU, engagement data is presented against an anonymised cohort average. This is useful as the student can understand their performance against their class and derive motivation from achieving against their class. Furthermore, a staff member can more broadly understand the engagement pattern of their students, and how individual students may map onto this. There are however some drawbacks from this method; half of the cohort will inherently be below the average which can be a demotivating factor for some, and performance against a whole assumes that there is only one correct way to engage as determined by the majority.

An interview with a member of staff at UMCU specifically for this output highlighted some useful ways to provide context to data. One option was to allow for a learning analytics system that compares an individual's data with that of the previous year of students. Another suggestion is to allow for a student to compare themselves to a specific student also on their programme. Both of these suggestions however have encountered issues with data protection. Clearly, there is a need for providing context to data, however the way that this is achieved needs to be considered on an institution- and outcome-specific basis.

4. Data Sourcing and Storage

In addition to identifying what data to use and understanding how to process it, data handlers must ensure that data pathways are robust, and that data is stored responsibly both inside and outside of the system. Breaking this down in more detail, we **recommend that you consider:**

4.1 Mapping the data pathways fully

Interviews with staff at NTU for this output specifically detailed issues in data pathways that feed the Student Dashboard systems on a daily basis. Data that is used in the learning analytics system is pulled from different systems from across the institution. A change made outside of the Dashboard system can still result in serious issues for data within the platform. If one imagines data as a river, and the learning analytics platform as a dam downstream, there may be sudden changes to the river several miles upstream that can affect what happens at the dam. Identifying this problem can become difficult without a clear understanding of where the data comes from.

An example of issues in the data pathway in real life was described during the interviews. One school (or faculty) within NTU made a subtle change to the module names on their own system, changing a 'full stop' to a 'comma'. When the data was pulled through to the Student Dashboard system, module names were no longer recognised and therefore the data feed failed completely.

Data pathways must also be considered after processing, as well as before, which was explored in our research as part of the OFLA project. Using data as part of mid-term reviews for example, highlighted a complex pathway of post-processed data in order for it to be used as an alert at key times during the academic year (NTU O9, 'Mid-term Reviews', section 2.2). Mapping the data pathway from source, to processing, to the user, is fundamental in preventing data issues and ensuring a smooth process.

4.2 Amending the original data sources before using

Along with considering data pathways, one must also consider data format and collection at its source. Learning analytics platforms in a university setting must pull data from different systems and areas within the university. As this data is predominantly internally generated however, there is an opportunity to amend the data at source, in order to ensure a more stable and more efficient system. There are several ways in which this can be achieved.

At UMCU for example, a review of the data used within the system as a pilot led to a change in the data format. Initially, student data that was selected to be used as part of a learning analytics platform was stored as a 'pdf' in a different learning management system. As this was complicated and difficult to subsequently be used as part of a learning analytics system, the data was changed at source to a different format in order for it to be pulled through more easily (UMCU, O9, section 5).

At NTU, staff in interviews conducted specifically for this output described how attendance data prior to its inclusion in the Student Dashboard was inconsistent across schools within the university. This lead to a review of how data was to be collected in some areas, with the introduction of QR codes to act as registers. This change in approach led to more accurate data, as well as a more efficient and timely way of collected attendance records in comparison to manual hand-recorded registers.

At AHS, researchers turned towards a greater use of student feedback data through self-completed surveys as a form of data (AHS O9, section 2.3). Introducing a data point that is controlled directly by the team who manipulate and process the data results in more flexibility in changing or amending that data as part of the support system.

4.3 Where data is stored and/or amended on your system

As described earlier in this output, learning analytics data is most valuable when it can be accessed by numerous stakeholders at their convenience. This means that data not only needs to be collated and processed, but must be held for an extended period of time in a secured system. There are therefore considerations regarding data protection, what data can be held, who can access this information, and how a student can access and change their own information if requested. This particular issue links to various points previously mentioned within this output, such as using graduated student data for contextual purposes, or whether data can be held by an organisation outside of the university itself, if an institution chooses to make use of an external provider (discussed later in this output).

The issue of storage however becomes more complicated with regards to sensitive information, or information recorded as part of the support process. Recording information virtually as part of the support process is key at all three OFLA institutions, with it either being current process (AHS O6, section 4.4; NTU O6, section 2.1.4) or planned as part of future updates (UMCU O6, section 2.4.2). Through our research, findings also suggest that students often present sensitive issues during support sessions, such as mental health concerns (AHS O12, section 5.3.3; NTU O6, section 3.4.3; UMCU O6, section 2.4.4). Recording information about a student that is sensitive is crucial to the supportive process, however with access to information on a learning analytics platform being open to staff and students, there are clear difficulties in recording confidential information. Choosing to not record this, or recording information in a vague or coded way can lead to students needing to repeat sensitive information to multiple staff members as they progress through the support process, which can cause significant frustration for the student (NTU O12, section 3.2). To what level this information is recorded, where, and who has access, requires deliberation.

5. Timing and the Academic Cycle

Research conducted as part of the OFLA project throughout has highlighted the importance of **when** data needs to be used, in addition to **what** the data is itself. An intervention must take place within a short space of time following the student experiencing an issue; if an alert is generated a significant time after the issue, the negative effects of that issue are likely to grow exponentially for the student. Support must also come at a point in the year in which the student has time to successfully take action and get back on track. Breaking this down in more detail, we **recommend that you consider:**

5.1 The advantages of 'live' data vs a 'period review'

The three OFLA institutions all take slightly different approaches with regards to their reliance on live data and exploring data at 'key points' in the year. At NTU, the Student Dashboard gives daily engagement ratings that reflect daily performance of the student, and alerts can be generated throughout the term. Staff understand that whilst this data is only a proxy for engagement, and that in most instances the data reflects what they observe in the classroom, an alert can still be key in situations where a student experiences a sudden and profound issue (NTU O6, section 3.2.2). This approach however relies on regular review of the data by staff, good data literacy to understand and interpret the data, and a consideration for context throughout the year, such as the student study cycle and whether the student can enact change to respond.

At AHS and UMCU, learning analytics data is more focused on periodic data. AHS have structured their use of big data around the review of FIT test results and attendance at a specific class (AHS O9, section 2.3). This delivers an alert for students at risk at a specific time of the year, and as a result, can include more structured support from the communication stage onwards. Similarly, UMCU use completion information and student feedback relating to the reasons for delays, which is available at key points in the year in order to understand student progress (UMCU O12, section 2.3). In both of these instances, the collation of data for period review gives greater flexibility in the data that can be used. By understanding context and expectations in advance, these institutions are able to create a support process based around data that is more informative at a specific point in the year.

Both of these approaches have advantages and disadvantages, and deciding the approach in advance allows an institution to make a more informed decision of what data is to be used as part of a learning analytics system.

5.2 When an alert is generated throughout a year…

Although at NTU alerts are generated throughout the term, the value of an alert being received can differ at various points throughout the year. At the start of the academic year, a student is more likely to be able to 'catch up' after overcoming an issue that has prevented them from fully engaging with their work. In the 2020-21 academic year, students were required to adapt to a blended learning approach and with the introduction of a second 'lockdown' as a result of the Covid-19 pandemic, were more likely to experience barriers to engagement. As a result, a calling campaign was introduced and focused on supporting students with low engagement at this key time in the year (NTU O12, 'call service campaign', section 2.1). This therefore required data to be available and processed in a timely way.

In addition to a call centre campaign, specific schools have identified key times within a term that students may need additional help. Reviewing data at a key stage in the year reduces the need for considering a context of where the student is in the study cycle, allows for staff training to be scheduled in preparation for the review of the data, and support can be structured and planned more in advance with a consideration for what is needed for the remainder of term. At NTU, a main example of using periodic data is as part of mid-term reviews, and a full critique of this process is detailed in an OFLA output (NTU O9, 'Mid-term Reviews'). At UMCU, students receiving low scores specifically for their first project are associated with subsequent failure or withdrawal, and therefore the collation and review of this data in the year is crucially important in supporting students (UMCU, O9, section 4.5).

Understanding when in the year an alert is most valuable is therefore critical in considering when to collect the data needed in student support.

5.3 …and when an alert is generated for the student

Finally, in addition to understanding when an alert is most helpfully generated throughout the year, OFLA research also highlights how different groups of students require different timings for alert generation. At NTU, an analysis of how much engagement data (or lack of any registered engagement over a continued period) revealed that an intervention is more likely to be successful when an alert is generated based on different time periods for first year students and final-year students. As a first-year student is more likely to participate in structured or scheduled teaching, a shorter timeframe of daily data can reveal issues for a student, in contrast to a final-year student, who is more likely to study independently and therefore require a longer threshold before an alert is generated (NTU O9, 'reducing the alert timeframe', section 3.2). There is an understanding here that for first year students, a small amount of engagement data is more accurate at predicting future success; this suggests that data handlers must consider the efficacy of their data and consider when alerts are generated accordingly.

Research conducted at AHS also indicated that understanding the type of issue that a student is experiencing, can also change how quickly a student acts to address the issue and how promptly the institution is required to support the student successfully. Psychology needs for example can lead to a student delaying seeking help, and therefore an institution must engage students with these types of problems more quickly, suggesting a need for more prompt alerting for these students (AHS, O12, section 5.3.3).

6. Stakeholders and Data Users

The OFLA project team consists primarily of data researchers, professional services staff, teaching staff, and study advisors. The full project team for learning analytics platform and/ or the use of data as part of the support process, however, requires staff from across the institution. Furthermore, staff from outside of the institution may be required to develop and maintain a platform. Finally, the inclusion of data users (such as staff and students) in the development of data or data platform is crucial in insuring it works correctly and remains effective. Breaking this down in more detail, we **recommend that you consider:**

6.1 Who you may need on a project team

As part of this output, staff at each OFLA institution were asked about who is involved in determining how, where, when and why data is collected and processed to support students.

Researchers in learning analytics are of course key in understanding why different forms of data are used, determining the outcomes following data review, and have a responsibility for the ethical treatment of user data throughout the process. Several other staff members however are needed to ensure this is conducted in the correct way. **Legal and compliance** staff for example can support this latter point, and are best placed to advise whether the data itself can be used from a data protection perspective.

Information Services staff (or whatever equivalent exists at your institution) are needed to help establish where the data is obtained, are fundamental in data management, and may be needed to provide ongoing technical support for any ongoing systems and processes. In addition to management of the system staff are needed for initial **project management and ongoing support** to administer the process from the top down as well as the bottom up. A system and its processes must also be embedded within the **policy and governance** of an institution (discussed further in O14), and therefore senior management are required.

Representation from each school (or faculty) is needed throughout not only to embed and manage the process at a local level, but to ensure specific data requirements are needed in areas that can differ at fundamental levels. In addition to this, **representation from users** is critical in ensuring data can be understood and used to its full extent, and that feedback can be acted on to ensure the process is refined.

Finally, **internal marketing and communication** staff need to be part of this process, for two main reasons; the first is to ensure users across the institution are aware of the process, and the second is to ensure essential training for users in using the data. For communication and training to be effective, this team needs to be part of the conversation from the start.

6.2 How to consult with the users

As noted throughout this output, the user must be at the centre of the system throughout the process of selecting, analysing, and presenting data, as well as the development of the system itself. This is not only to ensure data is understood, but that it continues to be relevant and applicable to the support process.

As part of the project, the OFLA team consulted users in the form of student surveys (NTU, O12, section 3.2), staff surveys (AHS, O12, section 3.2.1), to even including stakeholders in the design process (UMCU, O12, section 2.3). These are however somewhat ad hoc in nature and require the project researchers to develop methods of gathering feedback for a specific purpose. At NTU, the Student Transition Survey provides some opportunity to gather annual feedback from student users of the learning analytics system in order to make changes to the Student Dashboard (NTU, O6, section 4). At all institutions however, staff reflected on the need for an increased inclusion of stakeholder views, particularly in the form of co-creation and ongoing refinement of the process. Although this requires a significant amount of time and resource to embed, it is fundamental to a system that works well for the user as well as those staff members behind the data.

6.3 The benefits of internal vs external providers

In addition to involving internal staff from across the institution, staff external to the institution may be required in developing and maintaining the IT infrastructure and in data processing. The three OFLA institutions have approached this in differing ways, each bringing their own advantages and disadvantages.

At UMCU, the development of IT infrastructure and data processing is currently remaining 'in-house' (i.e. entirely utilising internal staff). An interview with a UMCU staff member as part of this output highlights some of the reasons for this. Firstly, developing new IT infrastructure in-house means that the institution retains complete control of the process. Secondly, sensitive data can be kept compliant with data protection legislation with ease when no data is being used or processed outside of the organisation. Thirdly, internal development of IT infrastructure can rely on internally agreed deadlines, that can be internally managed.

At AHS, the IT infrastructure remains an internal platform, however the data collected is determined in partnership with external organisations. The FIT test for example is developed in partnership with another external organisation, the University of Antwerp (AHS, O9, section 2.3). This has meant that the data itself is more robust, as it is built from a test developed using a larger more diverse cohort, thereby increasing the reliability of the data itself.

At NTU, the IT infrastructure itself has been designed and developed by a third-party provider, SolutionPath. An interview with a member of staff from NTU as part of this output highlights the benefits of using an external provider; that resources are available to provide continued developments to the system, that external providers have dedicated expertise to learning analytics systems, and that learning and research from other institutions can be applied to your own system. There are clear advantages to either an internal, external or mixed approach to the development and maintenance of IT systems and data, and these approaches need to be considered as part of this process.

7. Ongoing Development and Refinement

The majority of this guidance applies to the initial development of IT infrastructure and the early stages of managing institutional data. There are however continuing factors that an institution must consider on an ongoing basis. Support for data usage for example, from the system itself to user's data literacy, is often increasingly complex as time progresses. Maintaining trust in the system can easily degrade if and when issues arise, and without a firm understanding of the wider institution and its data processes, these issues are likely to arise more frequently than expected. Breaking this down in more detail, we **recommend that you consider:**

7.1 How you can prioritise ongoing support

Once the IT infrastructure is in place, and the data systems are set up to generate information on a regular basis, there is still likely to be the need for ongoing administrative and technical support. This is not simply to deal with problems as they arise, but often due to the inherent nature of data systems; infrastructure and processes that can be efficient and effective initially, can become burdensome over time merely as it accumulates more data. This effect was seen at NTU as part of the call campaign, whereby the excel spreadsheet that acted as a call database began to slow and become unwieldy as the term went on. This led to a longer-term recommendation of introducing a more robust CRM system in place of the current spreadsheet that had become overwhelmed (NTU, O12, section 4.2).

As part of the interviews for this output, a member of staff from NTU discussed this issue as part of the Student Dashboard. As a system encounters new and developing problems as a result of both external changes and the increasing amount of data, a significant amount of work is required not to develop the platform further, but simply to keep it functioning as intended. Particularly in the face of ever-changing context, from how students are able to engage with their studies to the indicators of success or failure to changing cohorts each year, a learning analytics system in particular has a greater chance of obsolescence or failure.

7.2 How you are consulted with institution-wide data system changes

Support platforms, whether they are learning analytics platforms or more general IT platforms, rely not only on data generated and collated by one specific research team but by a variety of teams across an institution. At AHS for example, the student support platform (SVS) relies on data produced by the FIT test, as well as personal data, enrolment data, and current academic performance gathered across that institution (AHS, O6, section 2). The Student Dashboard at NTU calculates engagement based on proxies such as library resources, attendance, and coursework submissions, all of which are generated by teams such as library services and teaching staff themselves (NTU, O6, section 2.1). The Osiris system at UMCU relies on data that is input manually, requiring staff themselves to input data in a specific way. In all of these processes, data used in student support is generated elsewhere in the institution, and changes to this data at any point can disrupt or even break the process entirely.

There is also data and information used in these systems that does not simply relate to the student directly that can be subject to change, that subsequently disrupts the system. An NTU staff member interviewed for this output specifically gave an example of one change that caused a failure in the Student Dashboard system entirely. This staff member described how each module is recorded centrally with a code, and due to a change in the way these codes are recorded ('full stops' in the code replaced with a 'comma'), the Dashboard system failed to recognise course codes and the system required a change in the basic programming to rectify the issue. This issue can be avoided if the team responsible for the learning analytics platform is made aware IT and data changes across the institution, that may have a knock-on effect on the platform itself.

7.3 How you can maintain trust in the data

In the event of data failures, or errors in the IT infrastructure, the immediate consequences are clear; correct data about students cannot be provided, and students may not get the support they may need at that time. There are however more long-term effects in the event of data issues or system failures. Through our research we have also heard first-hand from staff members who have described how confusion about learning analytics data has lead to distrust of the system, and even resentment that they have to use it at all (NTU, O6, section 3.2.3).

Interviews conducted specifically for this output highlighted how misconceptions about how the data is collated and presented can give the appearance of instability in the data, even if the process works as intended. An example given was of attendance data presented in the Student Dashboard at NTU. Attendance data in some areas is not immediately available, and must therefore be included retroactively at a later date; for the user this is shown as a change to historic data, and the user may see a difference in attendance for a specific day on difference dates. These changes having not been properly understood by the user then leads to distrust of the data, and therefore data handlers may also need to ensure that this is communicated to those that ultimately use the data itself.

7.4 How you can maintain data literacy and understanding

Finally, the issue of data literacy must be addressed specifically. This is an issue raised at points in this output and throughout research conducted as part of the OFLA project. Data handlers will have a developed understanding of the data they work with, however this knowledge must be conveyed to all staff both involved in developing policy, and the users themselves. OFLA research has found that when staff who do not understand the data, or believe that they can positively impact the data they are receiving, they are less likely to invest their time in the system itself (AHS, O12, section 3.3.1). This is more than simply consulting users about the system in its development, but ensuring **all users** have some level of data literacy and understanding about the data, and its usefulness. Again, students who do not understand the value of the data, are also much less likely to engage in the system, as found in the most recent Student Transition Survey.

Effective guidance is required to ensure staff usage, however even this guidance must be efficient and require little time to understand. Guidance and policy that is confusing or cumbersome has been found to still led to a lack of data literacy or understanding, as tutoring staff often do not have adequate time or resource to fully engage with the system and its complexities themselves (NTU, O6, section 3.5.3). Significant ongoing resource is required to improve data literacy amongst staff and students, and as experts, data handlers are an essential part of this process.

Concluding statement

Data and learning analytics are invaluable tools to support students at risk of failure. Although this project has identified strengths and issues within the process of using data, a number of these issues stem from the IT infrastructure and the system themselves. As is often the case in higher education institutions, staff are acting on data and systems that are already in place, adapting them to suit an increasingly important and complex function of student support. When introducing new IT infrastructure, or even in re-purposing existing data systems, lessons can be learned when reviewing the process in practice at the three OfLA institutions.

Although not an exhaustive list, this output produces a number of recommendations that are grouped under seven overarching themes. These seven themes are intended to help data practitioners consider their approaches to supporting early warning alerts. The sections provide a high-level series of considerations that may helpfully influence their thinking or approach. It is our hope that staff at higher education institutions can use these considerations to have a profound effect on data processes, before the first stage (alert) of the OfLA three stage model of support is even reached.