Higher Education (2025) 90:1049–1063
https://doi.org/10.1007/s10734-024-01363-2

Institutional constraints to higher education datafication: an English case study

Rachel Brooks[1] · **Jill Timms**[2]

Accepted: 13 November 2024 / Published online: 4 December 2024

Abstract
Strong claims are made by both higher education policymakers and institutional leaders about the advantages that can be brought about through 'datafication' (i.e. the quantification of human activities through digital means). Nevertheless, we know relatively little about how university cultures and processes may impinge on the rolling out of data practices, particularly in relation to administrative data (rather than that collected through learning analytics and other forms of user-generated data). This article draws on a particular case study—of attempts to widen participation to 'sandwich courses' in English higher education institutions—to argue that, in some areas of activity at least, institutional constraints serve to limit datafication substantially, including insufficient time and infrastructure, the prioritisation of data required for other purposes (typically national assessment exercises) and the role of institutional cultures. Cognisant of various problems associated with quantification, the article nevertheless contends that the barriers to the effective use of metrics identified in this research may have a direct impact on institutions' capacity to recognise and address social inequalities.

Keywords Data · Datafication · Higher education · Metrics · Sandwich courses · Widening participation

Introduction

Over recent years, both policymakers and university leaders have extolled the virtues of moving to a more metricised higher education sector: statistics about student satisfaction with their degree programme are held to improve the decision-making processes of prospective students, while data analytics are purported to help the shift to more personalised learning, for example (Gulson et al., 2022; Komljenovic et al., 2024; Selwyn & Gašević, 2020). In this article, we draw on a case study—of data use with respect to widening participation to undergraduate 'sandwich' courses (where students spend a year on a work placement, typically during the third year of a four-year degree programme)—to explore

✉ Rachel Brooks
rachel.brooks@education.ox.ac.uk

[1] University of Oxford, Oxford, UK

[2] University of Surrey, Guildford, UK

day-to-day data use in English higher education institutions (HEIs). In doing so, we argue that, despite the strong claims about the advantages of making more and better use of data, in this particular area of activity at least, significant constraints operate, limiting the advantages that can accrue through datafication. Indeed, despite our area of focus having strong connections to current policy priorities, in terms of both graduate 'employability' (which sandwich courses are commonly believed to facilitate) (Boden & Nedeva, 2010; Durazzi, 2021) and widening participation (Burke, 2020), we show how few HEIs in our research made systematic use of quantitative data in this sphere of activity. In explaining this apparent absence of datafication, we identify particular barriers to such data use—such as capacity constraints, the prioritisation of data required for other purposes (typically national assessment exercises) and the role of institutional cultures. We also show that while there was a clear absence of 'datafication' in this area across almost all institutions in our sample, there was also variation by HEI, seemingly linked to resource availability.

The article proceeds as follows: we first discuss the extant literature relating to datafication (with a primary focus on higher education). We then provide details about the study upon which the article is based, and the research methods we used, before outlining the relative absence of quotidian data practices and the likely reasons for this. We conclude by considering the implications of these themes for both scholarly understanding and HEI practice.

Background

Datafication, 'metric power' and higher education

Datafication is typically understood as the increasing importance of data in the everyday lives of people, and the drive to turn human behaviour and activity into data points that can be tracked and analysed (Meijas & Couldry, 2019). It is closely aligned with processes of quantification and the increasing use of metrics. Datafication has had a profound social impact—not least in the close links between the metrics generated through datafication and the operation of power. These are explicated well in the work of Beer (2016), through his concept of 'metric power'. He outlines four key characteristics of such power. First, he argues that it is based upon 'the creation and maintenance of limits and parameters' (p. 171). This is a form of control through limit, he suggests, by 'shaping edges to what can be known and by channelling activity in certain directions' (ibid.). Second, metric power is based on what is rendered visible and invisible. Some things are seen and so become important, while others are hidden and are thus deemed unimportant. Third, metrics exert power through their ordering, sorting and categorising—processes that are often transferred into the social world, creating divisions between people and facilitating patterns of judgement. Finally, metrics use particular models of the world that have the potential to become realities in their own right.

Higher education has, of course, not been immune from such influences (Burrows, 2012). Indeed, there has been a tendency within the sector to see datafication, alongside other social processes brought about by the increasing use of digital technology, as a 'revolutionary change' and 'a signifier of other ideas and values related to education – such as notions of freedom, speed and efficiency—which seductively give the impression that the digital can allow us to transcend the limits of the body and our social and material settings, or do away completely with notions of expertise or the need for teachers and so-called

traditional modes of scholarship' (Gourlay & Oliver, 2018, p. 1). Nevertheless, as Williamson et al. (2020) have argued, datafication in higher education (and in education more generally) needs to be theorised in different ways from datafication in other parts of life. For example, because it is typically understood as a public good rather than a commercial enterprise, the extraction of data from students and staff 'cannot be straightforwardly analysed as another instantiation of "surveillance capitalism", that is, the gathering of the "raw material" of human life *en masse* for analysis, sale and profit' (Williamson et al., 2020, p. 351). Instead, researchers need to be attentive to the distinct ways in which datafication may be enacted within the higher education context, in which data are rarely sold but 'metric power' is nevertheless frequently exerted.

The majority of the extant literature on datafication in higher education has, unsurprisingly, focussed on data collected about students (for example, through learning management systems, student surveys and management dashboards). Most commonly, researchers have examined the use and impact of learning analytics within higher education contexts. Thompson & Prinsloo (2023) provide a detailed list of what is commonly included in data analytics systems, and how this 'data frontier' is rapidly expanding to include new data sources:

> The data most often used in learning analytics are largely generated by institutional learning management systems …. Student information systems collect data about prior qualifications, socio-economic status, ethnic group, financial situation, and hours worked by students ….. Learners also create data traces though their digital artefacts and their movements may be tracked using geolocation data …. Emerging research suggests collecting biometric data from wearable devices to document students' stress and sleep patterns ….. More data creeps in as digital records are augmented with information about financial aid, disciplinary and criminal reports, personal health, and extensive data profiles from admissions applications. (p. 156)

Although some research has emphasised the benefits of datafication—for example, for overcoming the limitations of self-reported data from students about their own learning, providing personalised feedback to students about the next steps they should take to maximise their learning, and encouraging agentic behaviour (Selwyn & Gašević, 2020)—there is also a large body of work with a more critical orientation. This has emphasised, instead, problems with both underpinning assumptions and practical implementation. In relation to the former, Selwyn and Gašević (2020) have argued that the measures and proxies that are typically used in data analytics lack robustness, reliability and representativeness—because of the ambiguities and uncertainties bound up with processes of teaching and learning. A common further criticism is that learning is often a social and collective process, and yet data-driven approaches tend to frame learning in a highly individualised way, with data collected about the individual student, not the groups to which they belong (Gourlay, 2022; Selwyn & Gašević, 2020). Similarly, given the extensive body of evidence about the impact of social characteristics (such as class and ethnicity) on educational outcomes, scholars have contended that data-driven analytics and nudges cannot compensate for structural inequalities (Selwyn & Gašević, 2020; Whitman, 2020). In relation to practical implementation, research has highlighted the struggles lecturers experience in interpreting (often poor quality) data (Brown, 2020); the ways in which such 'data dashboards' can disrupt pedagogical approaches (Brown, 2020); concerns over surveillance, privacy and data exploitation (Gourlay, 2022; Kwet & Prinsloo, 2020; Marachi & Quill, 2020); and the propensity for correlations (between behaviours and outcomes) to be treated as causations (Whitman, 2020). Moreover, Thompson and Prinsloo

(2023) suggest that learning analytics exert a specific 'data gaze' (Beer, 2019), in which students (and staff) come to pass judgement on themselves according to how they believe they will be assessed by data analytics systems, irrespective of whether that accords with their own prior beliefs about learning.

Some scholars have explored students, not only as data subjects, but as active users of data. Reflecting on students' use of various higher education metrics even before they begin their degree course (for example, rankings and the UK website 'DiscoverUni'[1]), Komljenovic (2022) argues that such data position prospective students as 'calculative agents' entangled in processing of buying and selling. In this way, datafication, it is argued, can have a profound effect on student subjectivities. Others have pointed to students' capacity to resist some of the potentially negative consequences of datafication. Thompson and Prinsloo (2023), for example, suggest that students can resist and reshape the 'data gaze' in various ways—by uncovering how relevant algorithms work (how they consume, process, direct and represent data); how they can feed into the design and development of new technologies; how they can challenge invasive data collection; and how they can 're-story their data bodies when partialities and mis-representations become problematic' (p. 161).

A key theme of this broad body of work is the intractable nature of datafication (and associated quantification) in the ways in which HEIs operate. Chun and Sauder (2021), for example, show how, within the university in which they conducted research, the 'Evaluations Management Department' that was set up to collect and analyse performance data across the institution quickly assumed a significant degree of power. It was given authority to make demands of nearly every other unit in the university and reshaped ideas about what constituted good work: this became understood as work that helped the institution to perform better on various quantitative indicators. Similar assumptions about intractability are made frequently—sometimes explicitly or more commonly implicitly—in the literature.

National and international datafication drivers

Many of the practices within individual universities, outlined above, are informed by the wider national and international contexts within which they are located—and the perceived need to generate and/or prioritise data that will show the institution in the best light possible when made public (what we refer to as 'externally-facing data' in subsequent sections). With respect to staff data, for example, Pardo-Guerra (2022) has demonstrated how the UK's national research assessment exercise—the Research Excellence Framework (REF)—has had a profound effect on institutional practices as well as the subjectivities of individual members of staff, giving rise to what he calls the 'quantified scholar' (see also Burrows, 2012). Lim (2021) has made similar arguments with respect to the Danish higher education sector, demonstrating how the introduction of bibliometric research indicators by the government as a mechanism to allocate funding to universities altered staff perceptions and research practices—in what can be seen as a form of self-governance. England's national assessment of teaching—the Teaching Excellence Framework (TEF)—has had a comparable effect with respect to student-related data, as HEIs expend considerable effort in ensuring they perform well in the metrics that feed into this process, and pay less attention to aspects of student learning and experience that are not datafied in this way (Hayes

[1] DiscoverUni is a website run by the Office for Students for prospective higher education students. It provides comparable sets of information about full- and part-time undergraduate courses.

& Cheng, 2020; Kandiko Howson & Buckley, 2020). Espeland and Sauder (2016) have similarly shown how law school rankings, published regularly in the *US News and World Report*, have had a profound effect on law school practices throughout America, as every decision comes to be made with a view to its impact on the data that feed into the rankings.

Taking a more global perspective, Mills (2022) has examined the data infrastructures that, he contends, underpin the hegemony of Northern European university models. He identifies three specific groups of metrics that drive global higher education: statistical indicators (such as those published by the OECD in their annual 'Education at a Glance' publication), global rankings (notable examples are those published by QS and *Times Higher Education*) and citation indexes (including Scopus and the Web of Science). These are all typically structured around forms of knowledge and practices emanating from the Global North yet have influence across the world. Indeed, Mills (2022) writes:

> The easy "globalisability" of education indicators such as citation "impact" factors and rankings gives them discursive and policy power. The data produced by these infrastructures gets scrutinised and acted on by universities, governments and students. They lead universities to "see" globalisation as both an organising principle and a competitive marketplace. (p. 475)

As with forms of nationally collected data, such as the REF and TEF discussed above, the appeal of such metrics to policymakers and some university staff and students is that they seem to offer 'an information short-cut, effacing other genres of knowledge and expertise' (Mills, 2022, p. 486). Nevertheless, they frequently reinforce particular power relations, favouring certain agendas over others, imposing hierarchical relations between countries and/or institutions and conferring status to only a limited number of institutions (Decuypere & Landri, 2021; Pusser & Marginson, 2013). Datafication has also been driven by the actions of service providers and EdTech companies, working across national borders, that have sought to provide 'data solutions' to this growing market (Komljenovic & Robertson, 2016; Williamson et al., 2020). Reflecting on these trends, scholars have argued that datafication is evident in all parts of HEIs—part of a paradigm shift away from pursuing science and education as intrinsically worthwhile goods to focussing on metrics that link activities to multiple external audit exercises (Anonymous, 2019; Shore & Wright, 2024).

Despite this burgeoning literature on the datafication of higher education, there nevertheless remain some gaps. Indeed, within studies of HEIs, there has tended to be much less focus on the use of administrative data than that collected via other means, such as through learning analytics. We thus have relatively little detailed, empirical knowledge of how administrative data—relating to students, staff and institutional processes more generally—are collected, analysed and used in particular institutional settings, and how institutional cultures, norms and processes may impact on such data practices. In this article, we use one discrete area of activity (student participation in sandwich courses) as a lens to explore this specific facet of datafication.

Methods

We develop our arguments by drawing on data collected for a project on widening access to, and success on, sandwich courses. The project was funded by the Centre for Transforming Access and Student Outcomes in Higher Education (TASO) and ran from October 2022

until May 2023. A key reason for undertaking the research was because of national evidence that while sandwich courses can have a particularly beneficial effect on the employment outcomes of those from widening participation backgrounds (Kerrigan et al., 2018), such students are less likely than their peers to enrol on such courses (ibid.). We aimed to develop new knowledge about the barriers to enrolment faced by widening participation students and what could be done, across the sector, to improve access and ensure success for this group in relation to sandwich courses.

The project had various strands, including individual interviews with 40 undergraduate students from four HEIs, and the development of a Theory of Change (Brooks & Timms, 2023a, 2023b). However, in this article, we draw primarily on the interviews we conducted with staff members in ten English HEIs. We asked staff with responsibility for sandwich course provision a wide range of questions including about how such courses were organised in their institution, what they perceived to be the barriers to participation and success for students from traditionally under-represented backgrounds, and whether they had introduced any initiatives to support this particular group. In addition, we asked them a series of questions about the data they collected on those who undertook sandwich years to help us understand the extent and nature of any equity gaps in participation at the institutional level.

Ten HEIs were chosen to ensure diversity with respect to geographical location (within England), size and status, and history of offering sandwich courses (see Table 1). We included two from the 'Russell Group' of 'research intensive' universities; three 'newer' institutions, that received university status after 1992; and five 'older' universities outside the Russell Group. In each HEI, we interviewed at least one member of staff responsible for sandwich course provision. In most cases, such individuals held institution-wide roles, although in HEI 4, the interviewee held a faculty-level role (the majority of the institution's sandwich courses were, however, located in their faculty). In two HEIs (HEIs 6 and 8), two colleagues were interviewed together, as they both were keen to contribute to the research. Thus, in total, we interviewed 12 staff members. Each interview lasted about an hour and was conducted (and recorded) by the two authors using Microsoft Teams. Digital transcripts were generated. In addition, after each interview, the relevant researcher made notes of key points, using an analytical grid. An inductive analysis was then undertaken, through coding of the analytical notes and the interview transcripts. Codes were derived,

Table 1 Details of the HEIs at which staff interviewees were based

Reference number	Relative size of sandwich course programme	Type of HEI
HEI 1	Large programme (about 60% of students take sandwich course)	Pre-92, not Russell Group
HEI 2	Large programme (about 30% of students)	Pre-92, not Russell Group
HEI 3	Large programme (about 65% of students)	Pre-92, not Russell Group
HEI 4	Small programme (number for whole HEP not available)	Post-92
HEI 5	Small programme (under 1% of students)	Russell Group
HEI 6	Small programme (percentage not known)	Post-92
HEI 7	Medium size programme (about 9% of students)	Russell Group
HEI 8	Large programme (more than 50% of students)	Post-92
HEI 9	Small programme (4–5% of students)	Pre-92, not Russell Group
HEI 10	Small programme (4–5% of students)	Pre-92, not Russell Group

inductively, from the data and deductively, from relevant theory. It should be noted that the absence of data about sandwich courses, or lack of use of the available data, was not something we, or the sponsors of the research (who work closely with higher education providers), had anticipated prior to the start of data collection. Thus, theory in the area of datafication had not been used to frame the study at the start; rather, we turned to it during the process of analysis to help explain some of the emerging, unexpected empirical findings.

In the sections that follow, we focus on the ways in which data were (or were not) used by the individuals we spoke to and/or their institutions—one of the key themes to emerge from our analysis. In the interviews, this was discussed with respect to participation on sandwich courses, and the subsequent destinations of those who took a placement year. In both cases, we were interested in what kind of data was collected and how it was used—particularly in relation to tracking the experiences of students from widening participation backgrounds.

Significant constraints to datafication

As outlined above, a dominant theme in the recent literature on higher education is the 'intractable' nature of quantification within HEIs (Anonymous, 2019; Hillebrant & Huber, 2020; Williamson et al., 2020), and the ways in which such datafication is continually reshaping organisational structures, practices, power and culture (Chun & Sauder, 2021; Epseland and Sauder, 2016; Shore & Wright, 2024). In engaging with this work, we examine—with respect to one particular area of HEI activity—whether data are collected or not, whether they are appropriate for the purpose(s) they are intended to serve, and the extent to which they are used by relevant staff members. Given the focus of our project on widening participation to sandwich courses, and ensuring successful outcomes for all students participating in such courses, our assumption (shared by most of our interviewees) was that 'relevant' data would include demographic information, pertaining to widening participation categories, about those who had considered and/or taken up a sandwich course, and in relation to the outcomes (academic and employment-related) of sandwich courses. This would likely be derived from administrative sources and, possibly, student surveys.

While different types of data were referred to by most of our interviewees, and most thought that data could offer useful insights into the effectiveness of their area of activity, there was little evidence of the kind of 'intractable' data use discussed above. This was for three main reasons. First, in some cases, interviewees explained that no relevant data were collected. In relation to *access* to sandwich courses, four HEIs did not collect any data on this at all (HEIs 1, 4, 6, 8). With respect to the use of data on the *outcomes* of students taking sandwich courses, three HEIs (1, 4 and 8) did not collect or analyse data on this either. Moreover, although the HEI 10 interviewee reported having collected their own data on outcomes, they explained that this had been a one-off occurrence, and so it was not possible to map trends over time.

Second, in some HEIs, relevant data were collected but not analysed. In the following two quotations, for example, interviewees note that appropriate data—about students who take sandwich courses—are collected, but indicate that no systematic analysis of the data had taken place:

> We do collect statistics about those who take placements [sandwich courses], but I don't know what they show. (HEI 10)
> Interviewer: Are you aware of any disparities between groups of students in access to sandwich courses?

Interviewee: No, I think I'd have to look at the data. (HEI 5).

Such evidence tends to support the contention, made in a recent study by Komljenovic et al. (2024) that 'data lakes' are emerging, as HEIs collect more and more data, which, for various reasons, remain untapped.

Third, in other cases, appropriate data were collected and analysed, but in a very limited manner. For example, the interviewee from HEI 3 explained how data were collected and analysed in relation to the participation of students from under-represented ethnic groups, but not with respect to any other widening participation categories. This limited form of datafication, in which only some social characteristics were datafied, was not, therefore, able to inform any action with respect to the participation of widening participation students generally.

Indeed, across all ten HEIs, there was only one example where data were used in a systematic fashion to help analyse who was accessing sandwich courses within the institution, and the extent to which they were representative of the wider student population. The staff member from this institution (HEI 7) explained that they collected statistics annually—in relation to students who register as wanting to do a sandwich course placement and those who do ultimately go on placement. Data about their social characteristics are drawn from the HEI's central student record system, which facilitates analysis by specific widening participation category—particularly disability, ethnic background, socio-economic status and whether they come from a low participation neighbourhood.

Evidence such as this raises significant questions about the extent to which the HEIs involved in this research can be seen as 'datafied' spaces—i.e. spaces in which many aspects of activity are turned into appropriate data, and such information is accorded high-value (shown not least through systematic analysis and use). While scholars have argued that the effects of datafication and quantification can be seen at all levels of the higher education system, from policymakers to students and staff (e.g. Hillebrandt & Huber, 2020), our research suggests that there are some places, at least, where data use remains very limited—even when it would seemingly help institutions to assess their performance in strategically important areas (such as the opportunities available to students from widening participation backgrounds). Here, there are commonalities with Hillebrandt's (2020) study of German universities, in which he argues that despite a 'frenzy of datafication' across the sector, the impact on many day-to-day practices has remained relatively limited. In the sections that follow, we examine some of the factors that proved an impediment to the more widespread collection and/or systematic analysis and use of relevant data.

Capacity constraints: infrastructure and time

The most common constraint, identified by participants, was the lack of institutional capacity to collect and/or analyse appropriate data. Indeed, some interviewees noted that although there had been some attempts in their organisation to collect data to help understand which students were accessing sandwich courses and which were not, staff were often hampered by poor infrastructure. For example, the staff member from HEI 9 commented that they did not have a very good data system for placements—'we are still quite Excel- [spreadsheet] based'. Excel spreadsheets were viewed as limited as they could not be easily shared or updated, and data were relatively hard to manipulate. This, according to the interviewee, made collection of appropriate data laborious and systematic analysis of the data relatively difficult. They went on to reflect that they were able to access data

on specific characteristics if they needed them (giving the example of gender) but this was often a cumbersome process that did not incentivise regular engagement with such data.

Interviewees also pointed to the limited time staff had available to analyse data that the institution had collected. For example, the same interviewee (from HEI 3) who explained that their team relied heavily on Excel spreadsheets, with all the constraints that imposed, also noted that relevant data about students (including their demographics and whether they met any widening participation criteria) was held centrally in their institution, and in a form more manipulable than on an Excel spreadsheet. However, the centrally located staff who analysed these data did not have the capacity to do so for all areas of the organisation, and access to and performance on sandwich years was not deemed a high enough priority to warrant analysis.

Indeed, several interviewees described how 'externally-facing data'—i.e. that required by regulatory bodies and/or that fed into national and international league tables—was commonly prioritised, leaving little time for information officers to devote to generating and/or analysing data for internal purposes. The interviewee from HEI 1, for example, was unclear about what data, if any, were collected about equity gaps but believed that they were generally only pulled together for high-level reports 'such as for the TEF'. As discussed earlier in this article, it is argued that much of the datafication of higher education, in the UK at least, has been driven by various national imperatives, such as the Teaching Excellence Framework and the Research Excellence Framework, alongside global influences including international university league tables (Hayes & Cheng, 2020; Mills, 2022). Our research points, similarly, to the strong influence of such forms of externally facing data. However, in several of our case study HEIs, this seemed to be *limiting* other forms of data collection and analysis, rather than driving faster and broader datafication.

Recourse to (limited) externally produced data

Publicly facing data are generated not just by HEIs but by other sector bodies, too. Indeed, some of our respondents contended that, given the limits on the data their institution was able to generate itself, they had turned to externally produced sources of data, in the hope that it would offer insights relevant to their work on sandwich courses. The source most commonly mentioned (by half of the HEIs), when asked about such data was Graduate Outcomes—data derived from the nationwide survey that is conducted annually by the UK's Higher Education Statistics Agency which asks graduates about their destination (in terms of employment and further learning) 15 months after graduation.

Nevertheless, despite this being the only data source that many interviewees drew on (with respect to the outcomes of a sandwich year), most deemed it problematic and/or noted that they ran only very limited analyses on it. The interviewee from HEI 10, for example, described Graduate Outcomes data as 'patchy' (meaning that their reach and reliability were both considered poor) and explained that, for this reason, their institution had not made much use of it. Other participants asserted that while they used it to compare the outcomes of those students who took a sandwich year with those of students who did not, it was rare for such data to be analysed internally in terms of social characteristic—which severely limited the conclusions that could be drawn about the impact on students from widening participation backgrounds in comparison to their peers. The interviewee from HEI 9, for example, commented:

> Graduate Outcomes data indicate placements are positively correlated with employment success, but it is hard to say whether this is down to the placement or just the type of people who take placements.

HEI 6 *had* conducted an analysis of Graduate Outcomes data by social characteristic, remarking that 'when Graduate Outcomes data is analysed by whether students took a placement, it shows that they are more likely to have a professional job and be earning a higher salary – irrespective of social characteristic'. However, the interviewee also noted that this was a one-off analysis (conducted by centrally located staff) requested by the sandwich course team. It had not been repeated—because of the capacity constraints discussed above—and this made it difficult to trace patterns over time. Thus, with respect to this particular data source, concerns about the quality of the data or an apparent lack of time and/or resource to run additional analyses again bring some of the constraints to higher education data use into sharp relief.

Negotiating institutional cultures

A further barrier to using data to analyse access to and outcomes of sandwich courses was perceived to be the wider culture of the institution and, in particular, its attitude to risk. For example, the interviewee from HEI 2 explained that the data collected in their institution was limited to two main variables—subject of study and fee status (whether home or international)—because of widespread aversion to risk:

> We don't do as much [data collection] as we would like to on that because of ongoing cautiousness at the university about how some of that data is used and how it's shared with different teams. So we don't have as much data on for example how our WP [widening participation] students compare and perform relative to the overall cohort. It's slight frustration because we'd like to do more of that. …. It's a bit like the same as when GDPR [General Data Protection Regulation] first came in – the university took very much a kind of ultra cautious, safety-first approach and therefore that data was kind of [limited to] certain gatekeepers.

We discuss this further in relation to institutional differentiation below. It is notable, however, that in this quotation, the interviewee implicitly contrasts their perspective—keen to have more data available—with that of 'the university', which was seen to have taken a much more cautious approach. Such tensions were alluded to in many other interviews, too. Indeed, numerous participants outlined the struggles they had faced in gaining access to relevant data, and in influencing decisions about what should be collected and what analyses should be run. This was particularly marked in half of the HEIs (2, 4, 6, 8 and 9). In HEI 9, for example, the staff member explained how they were keen to track participation in sandwich courses over time by social characteristics but that this had not been facilitated by the institution. In other HEIs (such as 4 and 6), interviewees spoke of having to 'request' particular analyses to be run (which could be turned down), leading to a fairly ad hoc and inefficient way of proceeding, and illustrating the relative lack of agency accorded to staff—typically occupying mid-level organisational roles—in accessing and manipulating data. (In such cases, we would contend, processes of datafication are limited if staff are unable to access and/or use centrally-held data.) Moreover, the interviewee from HEI 8 claimed that although they did not collect any specific data on sandwich course

participation, this was on their 'wish list' of things to be able to do, recognising that this would facilitate the identification of any equity gaps.

These data suggest that, in seeking to develop a better understanding of processes of datafication within HEIs, it is important to continue to pay attention to relationships between staff and how power can be exerted between different levels and functional groups within an institution. However, rather than conceptualising this primarily in terms of surveillance and responsibilisation (Gourlay, 2022; Kwet & Prinsloo, 2020; Marachi & Quill, 2020), we also need to consider issues related to data access and analysis. In the cases discussed here, there is limited evidence of staff agency not only in relation to the re-shaping of data practices (Thompson & Prinsloo, 2023), but also with respect to much more basic issues such as data access.

Discussion and conclusion

Institutional constraints to datafication

Examining a discrete set of activities within the UK higher education sector—those relating to sandwich courses—has provided a useful lens to examine quotidian practices with respect to the availability and use of data. Despite the strong emphasis on data by government bodies and HEI senior management teams (Mills, 2022; Pardo-Guerra, 2022) and the claims made about the 'intractability' of HEI data use in the academic literature (Hillebrandt, 2020; Williamson et al., 2020), our evidence suggests that datafication is perhaps not as widespread as some have claimed. Indeed, our research indicates that some areas of activity—even those linked to high-profile political and institutional priorities (in this case, employability and widening participation)—have remained largely untouched by 'intractable' datafication, with relevant data either not being collected or, where it is collected, not being made available to staff working in pertinent areas. Our argument is not that every member of university staff will or should use data in their everyday practices. Rather, in this specific case, many of our participants themselves thought that more and/or better access to relevant data would enable them to carry out their roles more effectively and, in particular, take more meaningful action to open up sandwich course opportunities to all.

We have shown how, in many of the HEIs in our study, quantitative data are rarely collected and/or used in a systematic manner to examine whether those who embark on sandwich courses are representative of the wider student body and whether outcomes from sandwich courses are similar for students irrespective of their social characteristics. Our focus on specific social practices—including the limits on data use imposed by, for example, capacity constraints, the prioritisation of data required for other purposes (typically national exercises) and the role of institutional cultures and relations between staff—thus contributes to the body of work that has argued that higher education and digital technology are mutually shaping (Selwyn, 2014). It also helps to counter the tendency, often manifest in education, to see digital technology as ushering in revolutionary change (Gourlay & Oliver, 2018). Moreover, it suggests that the 'data imaginaries' that are played out at some levels of HEIs are fractured and limited and may be at odds with the grander imaginaries at play elsewhere within the sector, about how datafication will improve society (Williamson, 2017).

Our evidence also points to some differentiation across the sector, indicating that data practices, and thus perhaps also data imaginaries, may differ across HEIs, as well as within them. For example, it was notable that the only HEI which conducted a systemic and

annual analysis of sandwich course data, with respect to social characteristics (HEI 7), was one of only two institutions in our sample from the Russell Group of 'research intensive' universities. Typically, such institutions have a larger resource base than other HEIs, and it is thus perhaps unsurprising that HEI 7 was able to conduct this activity while others were not. Similarly, the HEIs that appeared to make the least use of any quantitative data about sandwich courses (HEIs 4, 6 and 8) were the three 'post-92' HEIs in the sample. In these ways, the widely documented inequalities between UK HEIs may be informing data practices. It seems particularly problematic that the less prestigious HEIs in this sample, which are more likely than their peers to have a considerable proportion of students from widening participation backgrounds (Boliver et al., 2022; Jones, 2022), seemingly had fewer resources to draw upon to assess whether such students were adequately represented on such courses. More localised institutional 'data cultures' appear also to play into this mix, such as the 'risk aversity' mentioned by the interviewee from HEI 2.

It is important to note that our research focussed on only 10 HEIs, and one area of practice. It is thus possible (and indeed likely) that at other institutions and in other areas of activity, much greater use may be being made of data. However, this case study does show quite clearly that 'intractable' datafication is not evident in all HEI activities, including those which could potentially benefit considerably from greater use of quantitative indicators. Indeed, while national and even institutional rhetoric may emphasise the significant improvements made to the student experience by the widespread use of metrics, our evidence shows that everyday data practices can be very limited in particular parts of HEIs. Quantification can, of course, bring problems of its own, not least the ways in which inequalities can become reproduced through algorithms and inappropriate proxies (Selwyn & Gašević, 2020; Zembylas, 2023). Nevertheless, we would argue that metrics can be used effectively to surface persistent inequalities—in this case, with respect to participation in and success on sandwich courses.[2] The barriers to the effective use of metrics highlighted in this article may thus have a direct impact on institutions' capacity to recognise and address such inequalities.

Implications for widening participation

Alongside demonstrating some of the constraints to datafication in a particular area of HEI activity, this article also supports claims in the wider literature about how 'metric power' is being exercised through specific data practices. As noted above, Beer (2016) has contended that a crucial way in which such power is exerted is through what is rendered visible and invisible. This is played out explicitly in our research in the relative inattention paid to students' social characteristics in the data collected and/or the ways they were it was used. As we have outlined above, very few HEIs in our sample collected data in any systematic fashion on the social characteristics of those who took part in sandwich courses or analysed Graduate Outcomes data to explore whether the outcomes of such courses varied by student profile. In the small number of cases where such data *were* made available, this was often in an ad hoc manner, requiring specific requests to be made to colleagues in other parts of the organisation—thus preventing longitudinal analysis and impeding efficient working practices.

[2] This is one of the key recommendations for change, outlined in the final report from the project (Brooks and Timms, 2023a).

As a consequence, the extent to which students from widening participation backgrounds were accessing sandwich courses—and then succeeding on them—relative to their peers typically remained invisible. While the majority of our interviewees were able to speculate on the extent of any under-representation and/or poor experience, this was typically on the basis of anecdotal evidence and their own 'sense' of how inequalities were played out in this area. Although reflecting on professional experience is obviously important, many inequalities may not be visible to staff (for example, if a student chooses not to talk about their neurodiversity or first-in-family status), even if they have regular contact with those eligible to take a sandwich course. Moreover, given the status often accorded to quantitative data within the senior management teams of universities (Beetham et al., 2022), the lack of any statistical reporting about inequalities by social characteristic, as they pertain to sandwich courses, makes it highly likely that such issues will struggle to gain the attention of senior leaders. In addition, two interviewees (from HEIs 3 and 4) thought that one of the reasons data were not collected by social characteristic in their institution was because of the perceived diversity of the whole institution, with a relatively large proportion of students categorised as from widening participation backgrounds. While this may well be the case, it does not follow that such students will automatically be well represented on sandwich courses and thrive during a placement year.

This 'invisibility' has similarities with various other types of 'invisibility' discussed in extant research—notably, the failure to ask questions relating to some equity issues in national data collection exercises (Hayes & Cheng, 2020) and institutional decisions not to include data on students' social characteristics within learning analytics systems (Whitman, 2020). Whitman's (2020) US study demonstrates how the absence of data on students' social characteristics restricted the number of spaces within the institution available to explore the impact on learning of discrimination and other social processes that are linked to students' gender, race, ethnicity and class. Although in her research, this was because of a conscious decision on the part of the institution (because they wanted to collect data on only aspects that they believed were in students' control to change), we contend that a lack of conscious decision-making (with respect to collecting data about students' social characteristics) can have similar results. This 'invisibility' is also in line with the wider body of work on the inequalities experienced by students from widening participation backgrounds that persist in UK higher education, despite decades of policy intervention (e.g. Boliver et al., 2022).

To conclude, our research has generated new knowledge about some of the institutional constraints to datafication in an area accorded high policy priority—because of its links to both employability and widening participation. We suggest, therefore, that until HEIs identify and address where data are still being under-utilised, opportunities are being lost. Indeed, our research provides an example of how removing specific barriers to data collection, analysis and/or use may improve access to and success on sandwich courses for traditionally under-represented groups.

Acknowledgements We are grateful to TASO for funding the research that this article draws upon. In particular, we would like to thank Rain Sherlock for overseeing the project so well and her extensive feedback on the final project report. We are also grateful to Adeeba Ahmad and Frances Gow, who were part of the project team and played key roles in setting up and conducting the fieldwork. Finally, we are indebted to all the staff (and students in the wider project) who generously gave up their time to talk to us.

Funding This research was funded by the Centre for Transforming Access and Student Outcomes in Higher Education (TASO).

Declarations

Ethics approval The research upon which this article is based received ethical approval from the University of Surrey's Ethics Committee (reference FASS 22–23 003 EGA).

Conflict of interest The authors declare no competing interests.

References

Anonymous (2019) Correlation in the data university: Understanding and challenging targets-based performance-management in higher education, *Acme: An International E-Journal for Critical Geographies*, 18, 6, 1184.

Beer, D. (2016). *Metric power*. Palgrave Macmillan.

Beer, D. (2019). *The data gaze*. Sage.

Beetham, H., Collier, A., Czerniewicz, L., Lamb, B., Li, Y., Ross, J., Scott, A.-M., & Wilson, A. (2022). Surveillance practices, risks and responses in the post pandemic university. *Digital Culture and Education, 14*(1), 16–37.

Boden, R., & Nedeva, M. (2010). Employing discourse: Universities and graduate 'employability.' *Journal of Education Policy, 25*(1), 37–54.

Boliver, V., Gorard, S., & Siddiqu, N. (2022) Widening participation in the UK: The possibilities and the limits. In C. Callendar, W. Locke, & S. Marginson (Eds.) *Changing higher education for a changing world.* Bloomsbury.

Brooks, R., & Timms, J. (2023a) *Improving access to, and success on, sandwich courses for students from widening participation backgrounds.* TASO.

Brooks, R., & Timms, J. (2023b) Widening participation to sandwich courses: Temporal challenges, *Studies in Higher Education* (online access).

Brown, M. (2020). Seeing students at scale: How faculty in large lecture courses act upon learning analytics dashboard data. *Teaching in Higher Education, 25*(4), 384–400.

Burke, P. J. (2020) Access to and widening participation in higher education. In P. N. Teixeira, & J. C. Shin (Eds), *The international encyclopedia of higher education systems and institutions*. Springer.

Burrows, R. (2012). Living with the h-index? *Metric Assemblages in the Contemporary Academy, the Sociological Review, 60*(2), 355–372.

Chun, H., & Sauder, M. (2021). The logic of quantification: Institutionalizing numerical thinking. *Theory and Society, 51*, 335–370.

Decuypere, M., & Landri, P. (2021). Governing by visual shapes: University rankings, digital education platforms and cosmologies of higher education. *Critical Studies in Education, 62*(1), 17–33.

Durazzi, N. (2021). Opening universities' doors for business? *Marketization, the Search for Differentiation and Employability in England, Journal of Social Policy, 50*(2), 386–405.

Espeland, W. and Sauder, M. (2016) *Engines of anxiety: Academic rankings, reputation and accountability* New York, Russell Sage.

Francis, P., Broughan, C., Foster, C., & Wilson, C. (2020). Thinking critically about learning analytics, student outcomes, and equity of attainment. *Assessment & Evaluation in Higher Education, 45*(6), 811–821.

Gourlay, L. (2022) Surveillance and datafication in higher education: Documentation of the human, *Post-digital Science and Education* (advance online publication).

Gourlay, L., & Oliver, M. (2018). *Student engagement in the digital university: Sociomaterial assemblages*. Routledge.

Gulson, K., Sellar, S., & Webb, P. T. (2022). *Algorithms of education*. University of Minnesota Press.

Hayes, A., & Cheng, J. (2020). Datafication of epistemic equality: Advancing understandings of teaching excellence beyond benchmarked performativity. *Teaching in Higher Education, 25*(4), 493–509.
Hillebrandt, M. (2020). Keeping one's shiny Mercedes in the garage: Why higher education quantification never really took off in Germany. *Politics and Governance, 8*(2), 48–57.
Hillebrant, M., & Huber, M. (2020). Editorial: Quantifying higher education: Governing universities and academics by numbers. *Politics and Governance, 8*(2), 1–5.
Jones, S. (2022) *Universities under fire.* Cham, Palgrave.
Kandiko Howson, C., & Buckley, A. (2020). Quantifying learning: Measuring student outcomes in higher education in England. *Politics and Governance, 8*(2), 6–14.
Kerrigan, M., Manktelow, A., & Simmons, E. (2018). Sandwich placements: Negating the socio-economic effect on graduate prospects. *Widening Participation and Lifelong Learning, 20*(4), 81–107.
Komljenovic, J. (2022). The future of value in digitalised higher education: Why data privacy should not be our biggest concern. *Higher Education, 83*, 119–135.
Komljenovic, J., & Robertson, S. L. (2016). The dynamics of 'market-making' in higher education. *Journal of Education Policy, 31*(5), 622–636.
Komljenovic, J., Sellar, S. and Birch, K. (2024) Turning universities into data-driven organisations: Seven dimensions of change, *Higher Education* (advance online publication) .
Kwet, M., & Prinsloo, P. (2020). The 'smart' classroom: A new frontier in the age of the smart university. *Teaching in Higher Education, 25*(4), 510–526.
Lim, M. (2021). Governing higher education: The PURE data system and the management of the bibliometric self. *Higher Education Policy, 34*, 238–253.
Marachi, R., & Quill, L. (2020). The case of Canvas: Longitudinal datafication through learning management systems. *Teaching in Higher Education, 25*(4), 418–434.
Meijas, U., & Couldry, N. (2019). Datafication, *Internet*. *Policy Review, 8*, 4.
Mills, D. (2022). Decolonial perspectives on global higher education: Disassembling data infrastructures, reassembling the field. *Oxford Review of Education, 48*(4), 474–491.
Pardo-Guerra, J. P. (2022). *The quantified scholar*. Columbia University Press.
Pusser, B., & Marginson, S. (2013). University rankings in critical perspective. *The Journal of Higher Education, 84*(4), 544–568.
Selwyn, N. (2014). *Digital technology and the contemporary university*. Routledge.
Selwyn, N., & Gašević, D. (2020). The datafication of higher education: Discussing the promises and problems. *Teaching in Higher Education, 25*(4), 527–540.
Shore, C., & Wright, S. (2024). *Audit culture*. Pluto Press.
Thompson, L. T., & Prinsloo, P. (2023). Returning the data gaze in higher education. *Learning, Media and Technology, 48*(1), 153–165.
Whitman, M. (2020). 'We call that a behaviour': The making of institutional data. *Big Data and Society, 7*(1), 1–13.
Williamson, B. (2017). *Big data in education: The digital future of learning, policy and practice*. Sage.
Williamson, B., Bayne, S., & Shay, S. (2020). The datafication of teaching in higher education: Critical issues and perspectives. *Teaching in Higher Education, 25*(4), 351–365.
Zembylas, M. (2023). A decolonial approach to AI in higher education teaching and learning: Strategies for undoing the ethics of digital neocolonialism. *Learning, Media and Technology, 48*(1), 25–37.