THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

The hidden architecture of higher education

OPEN ACCESS

UOC Universitat Oberta de Catalunya   Universidad de **los Andes** Colombia

RESEARCH ARTICLE

Open Access

CrossMark

The hidden architecture of higher education: building a big data infrastructure for the 'smarter university'

Ben Williamson (ORCID)

Correspondence: ben.williamson@stir.ac.uk
Faculty of Social Sciences, University of Stirling, Stirling FK9 4LA, UK

Abstract

Universities are increasingly organized and managed through digital data. The collection, processing and dissemination of Higher Education data is enabled by complex new data infrastructures that include both human and nonhuman actors, all framed by political, economic and social contingencies. HE data infrastructures need to be seen not just as technical programs but as practical relays of political objectives to reform the sector. This article focuses on a major active data infrastructure project in Higher Education in the United Kingdom. It examines the sociotechnical networks of organizations, software programs, standards, dashboards and visual analytics technologies that constitute the infrastructure, and how these technologies are fused to governmental imperatives of market reform. The analysis foregrounds how HE is being reimagined through the utopian ideal of the 'smarter university' while simultaneously being reformed through the political project of marketization.

Keywords: Dashboards, Data, Infrastructure, Marketization, Standards, Visualization

Introduction

Higher Education institutions have long collected large quantities of information about their students, programmes and facilities. With the emergence of new sources of digital 'big data' and systems to manage it, universities are increasingly being enmeshed in networks of digital data technologies and expert technical practices, and reimagined as 'smarter universities' (Lane & Finsel, 2014). The purpose of this article is to examine how complex data infrastructure projects within the HE sector are establishing a new hidden architecture of technologies, experts, standards, values and practices that will permit new big data technologies to be plugged-in to institutions and processes, in ways that may significantly transform HE itself. The core argument is that HE data infrastructures are not just technical programs but practical relays of policy objectives to reform the sector. In particular, data infrastructures simultaneously realize the utopian project of making smarter digital universities while also reshaping the HE sector through the political project of market reform. In this sense, data infrastructure constitutes a hidden architecture for marketization in Higher Education. The business of marketization in HE involves 'not just people, but technologies such as software, algorithms, computers, procedures and so on, in a rich collage of people, technology and programmes ... that align the work of the university with the logics of capitalist markets' (Komljenovic & Robertson, 2016: 633).

New forms of digitally generated data are already at the centre of many reformatory efforts in Higher Education at national and international scales. Digital data has been positioned by the UK government as a key element in a radical political reform to make HE more market-driven and customer-focused (BIS [Department for Business, Innovation, & Skills], 2011, 2016a). In this context, the UK government has begun considering the 'the role of big data and learning analytics for universities, including targeted marketing of prospective students, improving retention and personalising learning experience for individuals' (Westminster Higher Education Forum, 2017). The political think tank Policy Connect has produced the report *From Bricks to Clicks: The potential of data and analytics in Higher Education*, highlighting the use of 'fluid data' that is 'generated through the increasingly digital way a student interacts with their university' and 'learning analytics':

> Fluid data has the potential to provide an instant, accurate picture of how a student is performing—if it is able to be collected, linked and analysed. This is called learning analytics, which is the measurement, collection, analysis and reporting of data about learners for the purposes of understanding and optimising learning and the environments in which it occur. (Policy Connect, 2016: 4)

It concludes data 'has enormous potential to improve the student experience at university, by allowing the institution to provide targeted and personalised support and assistance to each student' (Policy Connect, 2016: 4). The government's 2017 Industrial Strategy announced £30million investment in 'innovative education technologies,' including artificial intelligence, to begin realizing this vision (HM Government, 2017).

Beyond the UK, the Society of Learning Analytics Research (SoLAR) has initiated the Support Higher Education to Integrate Learning Analytics (SHEILA) project to make universities more 'student data informed' and to build a Europe-wide policy development framework for learning analytics in HE 'that promotes formative assessment and personalised learning' (SoLAR, 2016). New international standards for learning analytics, to enable interoperability and data sharing across platforms and software packages, have been proposed by the EU Learning Analytics Community Exchange (LACE) project (Hoel, 2014). Similarly, the Learning Analytics in Australia project, funded by the Australian Government Office for Learning and Teaching, has sought to develop a framework for the advancement of learning analytics platforms and practices across HE institutions nationwide (Colvin et al., 2016).

Meanwhile, in the US the Institute of Higher Education Policy (IHEP) has initiated a Postsecondary Data Collaborative (PostsecData), with funding from the Bill and Melinda Gates Foundation, to advocate for 'high quality, robust and impactful' postsecondary data collection (http://www.ihep.org/postsecdata/about). Its aim is to establish a federal Student-Level Data Network (SLDN)—currently prohibited by US federal statute—that would 'leverage existing federal and institutional data to count all students and all outcomes to accurately represent the state of today's postsecondary system,' and would be hosted by a central statistical agency to 'align data across collections' and 'adapt best practices from existing data sharing efforts' (Roberson, Rorison, & Voight, 2017: 4). In addition, it has been claimed that 'to facilitate decision making and planning,' many US higher education institutions are 'creating dashboards to visualize data' and make it available and accessible to university managers and planners (Wolf, Taimurty,

Patel, & Meteyer, 2016). The think tank the Education Design Lab, which works with learning institutions, entrepreneurs and government on new models for post-secondary US education, notes that 'the dazzling possibilities of adaptive learning and big data have new companies scurrying to refine smart learning tools and algorithms, using predictive modeling to determine which students need interventions, and what kind of interventions are most likely to work' (Education Design Lab, 2014). Some 'horizon-scanning' studies even foresee artificial intelligence and 'natural user interfaces' that can be controlled through voice, gesture and facial expression as potential applications within HE, and argue that 'learning ecosystems must be agile enough to support the practices of the future' (Adams Becker et al., 2017: 2).

As these examples demonstrate, big data, interoperability standards, and educational analytics, dashboards and visualization methods are currently being promoted for use by HEIs at national and international scales. These efforts constitute a movement to create new big data infrastructures for the Higher Education sector. The collection, analysis and presentation of Higher Education data is being enabled by complex new data infrastructure systems that include both human and nonhuman actors: performance indicator metrics, data warehouses, data files, spreadsheets, information and records systems, visualization software, algorithm-led analytics packages and institutional dashboards, plus data managers, data stewards, business managers, financial officers and deans.

Building the data infrastructure of HE also requires private sector outsourcing companies, software developers, cloud hosting firms, data analytics designers and a host of other technical specialists, often working in organizations with global reach and activities that straddle the commercial and public sectors. Think tanks, consultancies and public-private collaborative alliances are providing discursive and evidence-based support for new data infrastructure projects. The data infrastructure of HE is therefore, like other infrastructures of transport, communication and power, the shared accomplishment of 'new constellations of international, inter-governmental and nongovernmental players' (Easterling, 2016: 15).

The UK Higher Education data infrastructure is one of the most developed in the world. Originally launched in 1994 in response to a government act that made the establishment of a coherent data framework for HE essential, it is currently the focus of a major redevelopment and upgrading project which is intended to use advances in information technology to enable 'richer data processes' and 'raise expectations about what data can do' (Youell, 2017). The rest of this article focuses in detail on the 'Data Futures' program begun in 2016 by the Higher Education Statistics Agency (HESA), the official statistics body for HE in the UK since 1993 (HESA, 2016). The Data Futures project is an integral part of a major governmental reform of the UK HE sector, the 2017 Higher Education and Research Act (HERA). Itself linked to the government's 2017 Industrial Strategy, HERA represents the culmination of a seven-year period of HE 'market reform' under the Conservative Party government which has seen massive cuts to public funding of universities, the introduction of student fees and loans, the entry of for-profit 'alternative providers,' and the escalating use of sector data to measure and evaluate institutional quality and performance (Burnett, 2017; Ridley, 2017). I examine several interrelated elements of the infrastructural constellation of Data Futures to indicate how it is constructing a new hidden architecture to enable big data techniques to become part of the HE environment following HERA. I make three key points: (1) HE data infrastructures are the products of policy networks of political actors, arms-length agencies,

consultancies, think tanks and private sector contractors whose reformatory objectives for the sector are encoded in the technical architecture. (2) Data standards that enable the infrastructure to function interoperably act as hidden rules to regulate how and which HE data are recorded and reported. (3) Data analytics visualizations and dashboards produced from information flowing in the infrastructure mediate how the sector is presented to publics, the media and policymakers—as well as how HE institutions view their own performance—and that visualizations act as sources for both institutional and political decision-making, action and intervention. Data Futures is not merely a technical project, but a political strategy to submerge and stretch a new infrastructure of data collection, processing and dissemination technologies and practices across the HE sector in order to enact market reform.

Datafication of higher education

Over the past two decades Higher Education in many countries has begun to experience a transformation as new forms of digital data are generated, analysed, and used to inform decision-making processes (Lane & Finsel, 2014; Mayer-Schonberger & Cukier, 2014). Many aspects of the university are now being reshaped by the production of digital data, from research and knowledge production, through institutional strategy and pedagogy, to state policymaking and governance (Selwyn, 2014). Since the early 2000s, new forms of 'data-intensive research' and 'data scholarship' have emerged across the sciences, social sciences and humanities, requiring universities to develop research capacity and expertise in collecting, creating, analysing, interpreting and managing diverse forms of digitized data (Borgman, 2015; Edwards et al., 2013). Synchronously, the increase in the use of market indicators, performance metrics, citation counts, impact measures, and the pressure of competition have affected academic work (Burrows, 2012; Lupton, 2015).

At institutional level, new forms of digital data enable administrators and managers to monitor organizational performance and improvement, as demonstrated by the rise of 'organizational analytics' and 'business intelligence' applications to support HE institutions (Guster & Brown, 2012). Moreover, data and its analysis are perceived to support responsive and effective learning, such as through the introduction of 'learning analytics' and 'adaptive learning platforms' that can sort and cluster student data and then 'feed-back' into pedagogic processes, influence the organization of curricular content, and 'personalize' students' learning experiences (Perrotta & Williamson, 2016; Wilson, Watson, Thompson, Drew, & Doyle, 2017). The emerging mobilization of big data-driven organizational and learning analytics in HE is part of a growing vision of the 'digital university' promoted by governments and businesses alike (Losh, 2014). The shared ambition to create 'smarter universities' is becoming the subject of feverish excitement—'institutions that can use the huge amounts of data they generate to improve the student learning experience, enhance the research enterprise, support effective community outreach, and advance the campus's infrastructure' (Lane & Finsel, 2014: 4). Student data in particular, it has been claimed, will 'reshape learning' through 'datafying the learning process' (Mayer-Schonberger & Cukier, 2014). The datafication of HE has previously materialized in the rapid growth of the commercial education technology sector (Selwyn, 2014), the expansion of the MOOC (massively open online course) across universities worldwide (Knox, 2017), and the evolution of methods and applications of learning analytics, educational data mining, intelligent tutoring systems and even artificial intelligence (Dillenbourg, 2016).

Quantitative big data have become significant for market-led HE policymaking too (Komljenovic & Robertson, 2016). Advocates for big data-driven reforms in HE are seeking to 'take data analytics to scale' so that education systems at national, state and district scales might 'take advantage of the massive amounts of data now being produced in ways no single campus can' in order to inform 'strategic decision making as well as development of more robust predictive models that can be used to improve student success' (Lane & Finsel, 2014: 17). From a more critical perspective, however, Selwyn (2014) notes that ideals of the digital university favoured by policymakers and policy influencers emphasize neoliberal logics of economics, efficiency, competition, audit, accounting, performance measurement, quality management, marketization, commercialization and privatization. Digital data systems in particular are used for surveillance and monitoring of staff and students, with 'dataveillance' techniques that produce 'administrative identities' from data for targeting and intervening on individuals (Selwyn, 2014). Moreover, in national contexts where HE has experienced market reform, student data has become an important source of evidence for institutions wishing to demonstrate their competitiveness (Ridley, 2017). Student data demonstrating institutional effectiveness in terms of outcomes and quality can therefore be used to attract prospective students, as business intelligence for internal decisionmaking, and as centralized sources for inspection by policymakers and politicians that can be used to evaluate institutional outcomes, create ranked league tables, and award or withhold financial resources (Burnett, 2017). These kinds of large-scale, national and state-level attempts to capitalize on huge volumes of institutional and student data crucially depend on major information infrastructure.

Following infrastructures

Infrastructures are 'the physical networks through which goods, ideas, waste, power, people, and finance are trafficked' (Larkin, 2013: 327), although as material forms they also require much work to build, maintain, repair and sustain. The analysis below develops an infrastructural approach informed by science and technology studies (STS) and related critical data studies. In such work, infrastructures are viewed not just as built structures or physical substrates which form a 'neutral background that enables an infinite set of activities' (Slota & Bowker, 2017: 530). Instead, infrastructure appears as a thoroughly heterogeneous and interpenetrating 'assemblage' of technological objects, standards, values, administrative procedures, and organizational work, all of which involve myriad people, institutions, technologies, policies, legalities, and financial arrangements to build, repair, maintain, and reconstruct (Kitchin & Lauriault, 2014). The 'infrastructures of daily life' are thus assembled from 'a maze of cables, connectors and infrastructural components' but also 'regulatory authorities who authorize interventions…, committees that resolve conflicting demands in the process of setting standards, governments that set policy, bureaucrats who implement it, marketers who shape our views of the role of the infrastructure in our lives, and more' (Dourish & Bell, 2011: 4–5).

Infrastructures also consist of sockets and switches which allow other technologies and practices to be 'plugged-in,' so infrastructures are scalable and mutable as they become 'networked' with other systems:

> Consequently, when a need arises to link heterogeneous systems into *networks*, devices, and/or social apparatuses known as *gateways* … must be created. The network phase signals not only the involvement of many more actors but also growing social

commitments manifested in, for example, explicit standards, user habituation, and organizational routines. (Plantin, Lagoze, Edwards, & Sandvig, 2016: 3)

The creation of networked gateways or sockets between heterogeneous systems within an infrastructure, however, often occurs incrementally as 'changes take time and negotiation, and adjustment with other aspects of the systems involved' (Bowker & Star, 2000: 35), and 'fully developed infrastructures are complex ecologies whose components must continually adapt to each other's ongoing change' (Plantin et al., 2016: 4). Consequently, infrastructural assemblages 'evolve and mutate as new ideas and knowledges emerge, technologies are invented, organisations change, business models are created, the political economy alters, regulations and laws are introduced and repealed, skill sets develop, debates take place, and markets grow or shrink' (Kitchin & Lauriault, 2014: 7). Notably, too, infrastructure studies have drawn attention to the utopianism of many projects. The materiality of an infrastructure is understood to carry political force as it acts as the material substrate of a certain kind of imagined utopian future (Larkin, 2013).

Infrastructures constructed to enable the collection, archiving and sharing of data via connected digital technologies have become part of everyday life over the last two decades. As such, a 'data infrastructure' is 'the institutional, physical and digital means for storing, sharing and consuming data across networked technologies' (Kitchin, 2014: 32). In particular, Kitchin (2014: 34) has highlighted how data infrastructures implemented by national statistics agencies often depend on interoperable software, shared services, analysis tools such as data visualizations, shared policies, and strong rules relating to data standardization, data quality and compliance, which 'enable data to be distributed, linked together and analysed.'

With infrastructure, then, there can be no simple distinctions of science, technology, values, society and political power, since they all interpenetrate one another and occur synchronously (Slota & Bowker, 2017). The purpose of this article is to survey and map the 'utopian' data infrastructure being catalysed by the HESA Data Futures program, and to consider how such an infrastructure may act as the sociotechnical materialization of particular social, political, technical and economic ambitions for the future of Higher Education.

Methodologically, the article reports on an initial attempt to 'follow the actors' involved in infrastructure-building, 'the formalizers, pigeonholders, categorizers, and number crunchers,' and to trace 'the making, the fine-tuning, the dissemination, and the upkeep' of infrastructure they perform (Latour, 2005: 227). As Kitchin (2014: 188-89) argues, there is a pressing need for case studies that trace out the sociotechnical arrangements of data infrastructures, especially genealogical studies of the situated unfolding of ideas, decisions, constraints, actions and actors that shape their development, evolution, influence, dead-ends and failures, as well as textual analyses of the discursive products that mediate their messages (writings, images, speeches, websites, brochures) and persuade people, companies and institutions to their logic. The initial analysis presented in this paper draws on a documentary analysis of materials produced by the HESA Data Futures program over its first 18 months, part of a planned longitudinal project to follow its infrastructural making, fine-tuning, dissemination, and upkeep. The data include project websites, published PowerPoint presentations, reports, white papers, and press releases, as well as governmental and consultancy documents that prefaced it in the

preceding few years, and the websites, documentary products and materials of partnering organizations involved in these projects. By following the sociotechnical, human and nonhuman arrangements involved in the early stages of Data Futures and tracing its textual mediation, the analysis emphasizes the dynamic networks of organizations, technologies and policies involved in data infrastructure-building, the task of standardization that underpins infrastructures, and the centrality of data analytics, dashboards and visualizations to communicating what data infrastructures are doing and producing. The discussion considers Data Futures as exemplifying how the new hidden architecture of marketization in Higher Education is the product of increasingly powerful nonhuman technologies and extra-state actors fusing together with the reformatory ambitions of political agencies of the state.

Infrastructural policy networks

Any infrastructure consists of complex networks of technologies, organizations and policies. This section details how HE data infrastructure upgrade in the UK is being accomplished by a cross-sector policy network of government departments, public agencies, consultancies, think tanks and software vendors, whose combined activities are building a technical system to deliver strategic political reforms. The redevelopment of the HE data infrastructure is, in fact, a practical enactment of the government's reformatory objectives to create a marketized HE sector.

A policy genealogy of data futures

The Higher Education Statistics Agency is the main national body for statistical data collection and analysis across the HE sector in the UK. A charitable company operating under a statutory framework on behalf of the funding councils and UK government departments, HESA's main remit is to support HE providers in fulfilling their data reporting requirements (https://www.hesa.ac.uk/about). HESA was formed in 1993 in response to a governmental working party on HE statistics and a subsequent act of parliament that gave it responsibility for building a system to collect HE data from 1994.

Funded by subscriptions from UK HE providers, HESA gathers information about all aspects of the UK HE landscape, including data on students, staff and graduates, finances and estates, academic departments and courses, and public engagement and commercial enterprises. One of HESA's key roles is to maintain UK 'performance indicators,' which provide comparative data and benchmarks on the measurable performance of HE providers across several areas in order to contribute to greater public accountability by the sector. Its data and analysis also enable strategic planning, inform policymaking, advance academic and commercial research, understand social and economic trends, and, finally, support prospective students' decision-making. HESA also, however, has a transformative role to upgrade the HE sector's technological infrastructure, a task it is undertaking through its Data Futures innovation program (https://www.hesa.ac.uk/innovation/data-futures). HESA's corporate strategy for 2016–2021 emphasizes its key objectives to upgrade the data infrastructure, improve sectoral data capability, and enhance insight through business intelligence, next-generation data analytics and visualization technologies (HESA, 2016). Launched in 2016, Data Futures is scheduled for live operationalization in 2020.

Data Futures has had a longer development period than its 2016 launch at first indicates. It was originally initiated in response to a 2011 white paper by the government Department of Business, Innovation and Skills (BIS) entitled *Students at the Heart of the System* (BIS, 2011) as part of a long-term government HE reform program later detailed in the 2016 white paper *Success as a Knowledge Economy* (BIS, 2016a). Together, these papers constitute a reformatory vision for UK HE as a whole that emphasizes student choice, a competitive marketplace of HE providers, increased performance measurement, improved outcomes, and future productivity for the economy. Two key reformatory recommendations of the white papers—that students in England should pay full fees for their degree courses, and the establishment of a Teaching Excellence Framework (TEF) to assess and rank university teaching quality—have already become the reality of HE in the UK. They are the subject of outspoken public debate on 'market reform' of the sector (Burnett, 2017; Ridley, 2017). HE reform was also a major part of the government's 2017 Industrial Strategy, which announced investment 'to test the use of AI and innovative education technology' in university courses to develop 'digital skills' (HM Government, 2017: 41). Underpinning these policy developments, however, is a less visible project to build a hidden architecture for the collection, analysis and dissemination of data required by the reforms. The white papers proposed specific reforms to the HE data and information landscape in order to arrive at a new system that could meet the needs of a wider group of users, reduce duplication, and result in timelier and more relevant data.

At the same time, the Department for Business, Innovation and Skills proposed a new governmental Office for Students (OfS). Described as 'explicitly pro-competition and pro-student choice' as well as a 'consumer focused market regulator' and 'non-departmental public body' operating 'at arm's length from Government' (BIS, 2016a), the OfS would act as a public regulatory body to explicitly champion 'the student, employer and taxpayer interest in ensuring value for their investment in higher education':

> Given the student is now the primary funder of higher education, there is a case for a new regulator that is capable of regulating the whole sector and operating on behalf of the student by supporting a competitive environment to promote choice, quality and value for money (BIS, 2016b).

With the legislative enactment of the Higher Education and Research Act (HERA) (http://www.legislation.gov.uk/ukpga/2017/29/part/1/enacted), the OfS announced its formal remit in summer 2017 with the ministerial appointment as chair of Sir Michael Barber, with a schedule to become operational in early 2018. The WonkHE think tank for HE policy named Barber—formerly a government 'delivery' adviser and later chief education adviser to the global education business Pearson—the most powerful person in UK HE in 2017, with a 'legendary fondness for metrics' (Leach, 2017). In an earlier report for the Institute for Public Policy Research, Barber argued that universities would need to reinvent themselves for an increasingly competitive marketplace of HE providers, using technology to aid in this transformation (Selwyn, 2014). Under Barber's leadership, a key responsibility of the OfS is assessing and rating the quality of, and the standards applied to, HE, and it has the duty to compile and make available higher education information along with regulatory powers to 'de-register' HE providers who fail to meet the designated standards. As commentators have pointed out, as a regulator

for HE following the enactment of HERA, the OfS represents both the increasing marketization of HE and the growth of a sector of commercial HE providers:

> The creation of a full-blown market for higher education with its own regulator—the Office for Students—heralded by the recent Higher Education and Research Act, is now nearly complete. The theory was that competition amongst the hundred or more existing universities, and a raft of new commercial providers, would hone the system to perfection. (Burnett, 2017).

In addition, according to the WonkHE think tank, it is widely anticipated by the sector that the OfS will replace periodic review based on annual data submission with a new method to use 'live data' and 'real-time metrics' to monitor institutions (Carrigan, 2017). Late in 2017, as part of the HERA consultative process, HESA was the sole submitting agency for the role of 'Designated Data Body' to work with the OfS to realize this ambition. It is in the context of the Industrial Strategy, the establishment of the OfS, the enactment of HERA, the designation of HESA as an official 'Data Body', and the political priorities they put on creating a marketized HE sector that a new data infrastructure has been proposed and developed.

Data inventories and blueprint models

In order to deliver the new data system required by the reforms, the multinational consulting firm Deloitte was commissioned in 2012 by the Regulatory Partnership Group (a collaboration between the UK's HE regulatory agencies) to 'produce a proposal for a coherent set of arrangements for the collection, sharing and dissemination of data for the higher education data and information landscape' (Deloitte, 2013). Subsequently, another global consultancy firm, KPMG, was appointed in 2014 to investigate and develop a blueprint for how student data should be collected by key stakeholders across the sector, as part of the Higher Education Data and Information Improvement Programme (HEDIIP) begun in 2013 (https://www.hediip.ac.uk/). HEDIIP was itself hosted by HESA—there is a traceable flow of staff and publications between HEDIIP and Data Futures—but retained its independence through oversight by a separate Programme Board, and involved a wide range of stakeholders from across the academic, government and industry sectors.

The outcomes of HEDIIP were reported to HESA by KPMG in 2015 (KPMG, 2015). HEDIIP had included an inventory of HE data collections which found a 'shocking level of duplication', reported 'a massive burden on the sector and silos of data that are not comparable', as well as 'no robust information on standards of data management and governance in institutions but plenty of anecdote to suggest that while some institutions are getting to grips with the challenges of understanding and managing their data assets, in others high levels of duplication and low levels of oversight and control are not uncommon' (Youell, 2015a). To tackle these problems, the KPMG HEDIIP report included specific blueprint proposals for a 'New Data Landscape', envisioned as 'a data and information landscape for Higher Education in the UK that has effective governance and leadership, promotes data standards, rationalises data flows and maximises the value of technology and enables improved data capability' (KPMG, 2015: 9). The report proposed the establishment

of an independent national 'Data Governance body' for HE to be based at HESA, which would then take responsibility for maintaining data standards and compiling a standardized dataset across the sector.

A 'theoretical target operating model' for HE was also envisaged in the KPMG report. This 'simplified vision of the future' of HE data infrastructure proposed a single central 'HE data warehouse,' based on cloud storage, whereby all student data would flow continuously between HE providers, uniquely-identifiable students and service providers and enable real-time analytics to be utilized (KPMG, 2015: 54). Though theoretical, this operating model underpins the proposed HEDIIP blueprint later developed as the Data Futures model. In the blueprint, HESA acts as a single centralized data collector and governance body for HE data, with other data collectors then accessing information via a HESA-maintained data warehouse, mediated through analytics functionality provided by other third-party data services providers.

The HEDIIP report also detailed a number of political, economic, social, technological, legal and environmental drivers and opportunities associated with this model. Politically, it would provide more timely and relevant data for governments and funders to make HE policy decisions; economically, it would offer cost-saving benefits; socially, it would reduce inefficiencies, streamline student management processes and enhance the perception of the sector; technically, it could deliver the benefits of data analytics and business intelligence insights; legally, it would not require complex legislation; and environmentally, it would reduce waste, inefficiency and the need for paper and manual processing (KPMG, 2015: 69).

A new 'data platform' for HE

The Data Futures programme itself commenced in 2016, with a four-year timeline for stakeholder consultation, system design, piloting, and operational rollout in 2020. According to official documentation, the cost of Data Futures was originally covered by a joint grant of £7.4million from the UK HE funding bodies (https://publications.parliament.uk/pa/cm201617/cmpublic/HigherEducationandResearch/memo/HERB30.pdf). Building on the blueprint and a detailed timetable of milestones provided by HEDIIP, Data Futures has the specified aim of transforming the HE data collection system that has remained largely unchanged since its inception in 1994, despite rapidly escalating demands for data from a range of organizations, such as the Office for National Statistics and incoming government bodies such as the Office for Students. A Data Futures PowerPoint presentation was made available on the HESA website in summer 2017 to promote greater sector awareness of the program:

> Data, whether produced, processed or consumed plays an essential role in understanding and supporting the development of the UK's HE sector. This data is used by potential students to make choices about their studies, and by government bodies to develop and review policies. HE providers need data to benchmark their operations, and to improve their efficiency and effectiveness, and the funding bodies use it to allocate public money. Data is also required for regulatory purposes, and in some cases is collected as a statutory requirement. (HESA, 2017)

In order to meet these needs, one of the main outputs of Data Futures is a planned 'Data Platform' to act as a hub for data collection and streamline demands on

individual Higher Education institutions by making HESA the main source for all sectoral data.

A technical specification for potential suppliers to build the platform was released by HESA in 2016 (https://www.hesa.ac.uk/files/Data-futures/Data-Futures_Overview-for-PQQ.pdf). The specification reveals the platform would include a vast number of interconnected technical components, including three 'user interfaces': a data collection portal, an analytics portal and a governance portal. Underlying these interface portals would be a range of 'services,' all underpinned by 'human and machine readable specifications,' a 'logical model' and 'physical data model,' a 'unique student identifier lookup service,' and a 'reporting engine.' The data platform would also include cloud storage, encryption, secure file transfer services, metadata, code, rules, data files, metrics, and specifications in terms of quality, reports, and data delivery, and more.

A key aspect of the proposed data platform is the involvement of partners from both the public and private sectors. The data collection elements of Data Futures have been outsourced for delivery by Civica, a global technology company which 'provides a wide range of software, digital solutions and technology-based outsourcing' for 'organisations to improve and automate the provision of efficient, high quality services, and to transform the way they work in response to a rapidly changing and increasingly digitalised environment' (https://www.civica.com/en-GB/what-we-do/). In addition to work in the commercial and financial sectors, Civica serves government and national security, health care, housing, local government, public safety and education sectors. Appointed by HESA in 2017, Civica Digital was contracted to work with the Data Futures team at HESA to develop a user-centred interface for HE providers to find more insights from their data in areas including student recruitment, retention and performance (Say, 2017).

In the earlier prototype stage of Data Futures, guided by the HEDIIP blueprint, much of the analysis being undertaken to deliver the new data landscape was enacted through a collaboration formed in 2015 between HESA and Jisc (Joint Information Services Committee) known as Analytics Labs. Jisc itself acts as the 'UK higher, further education and skills sectors' not-for-profit organisation for digital services and solutions,' and operates 'shared digital infrastructure and services' in order to 'deliver considerable collective digital advantage, financial savings and efficiencies for UK universities, colleges and learning providers' (https://www.jisc.ac.uk/about/who-we-are-and-what-we-do). The HESA/Jisc Analytics Lab collaboration is in many respects a prototype of the kind of data practices envisaged by Data Futures, emphasizing efficient data collection, removal of duplication, and capacity-building in cutting edge data analytics and visualization techniques.

Embedding 'big data' in HE

Significantly, Data Futures has been promoted by HESA's chief executive, Paul Clark, as a strategic program to embed 'big data' in UK Higher Education (Clark, 2015). Clark has described how the environment in which HE institutions operate is becoming increasingly data-intensive and data hungry, with policymakers, students and potential investors all seeking data and information for their own purposes and needs. At the same time as these 'trends are being driven by developments in higher education policy,' adds Clark, 'changes in the worlds of data, digital service delivery, and technology' are taking place as big data technologies and practices are embedded across sectors and industries. Late in 2017 HESA

organized the Data Matters conference to disseminate Data Futures updates to HE data practitioners, and as a forum for presentations and discussion around issues of big data, learning analytics, personalization services, data visualization, as well as data protection, privacy and data quality assurance (Guy, 2017). The Data Futures platform, then, is envisaged as a network with gateways and sockets for plugging-in other emerging technical innovations, which might enable it to become an infrastructure for live data analytics and real-time metrics. Consequently, claims Clark (2016):

> In ten years' time, it's possible to envisage a digital HE sector, with data-driven universities operating within a smart, connected environment. In this vision, universities would routinely use data drawn from many sources and devices to design and deliver their services, allocate resources, and monitor their performance. Policy-makers would similarly pool data from across government and the public sector to design interventions, monitor progress, and gain a far better and more granular understanding of how policy should be designed and delivered in order to achieve their aims. And users of the system would be able to access critical real-time data and information on their own progress, the resources available to them, and what they can do to maximise their chances of success.

These comments indicate how Data Futures is intended as a strategic intervention to make UK HE big data-ready, with standards and systems in place to allow institutional self-monitoring, to assist policymaking and intervention design, and to enable users to access real-time information on progress and available resources. The corporate strategy published by HESA in 2016 reinforces the 'transformative change' being wrought by big data in relation to higher education (HESA, 2016: 3). In other words, Data Futures is establishing the necessary infrastructure to support the application of learning analytics, adaptive learning platforms, and other forms of big data, AI and machine learning within digital courses, as current governmental interests in these technologies attest (HM Government, 2017; Policy Connect, 2016; Westminster Higher Education Forum, 2017).

At the core of this vision is a transformed view of HESA too, as an integral data warehouse for the collection, processing and dissemination of HE statistics, and potentially for new sources of big data. Through Data Futures and the new architecture for HE data it is building with its partners, HESA has been positioned as a new 'centre of calculation' (Latour, 1986) in Higher Education, a centralized data collector which is able to gather information from institutions distributed around the UK, transport those data to its servers and data processing centres, and then analyse, visualize and disseminate the data in order to shape knowledge and decision-making practices. HESA has ultimately been positioned to construct the hidden architecture of data collection necessary to the enactment of the 2017 Higher Education and Research Act.

In sum, as an infrastructure-building project, Data Futures is an accomplishment of a complex web of people and organizations, technologies, and social, political, legal and economic drivers. It operationalizes political, regulatory and economic demands on the HE sector, by standardizing and quantifying student data as a means toward ensuring that students', investors', and taxpayers' interests are adequately served. Data Futures is a practical enactment of politically-motivated market reforms to UK HE that centre on choice, competition, performance metrics, efficiency and accountability. As such Data

Futures needs to be understood infrastructurally as a dynamic sociotechnical assemblage of people, organizations, policies, technologies and legalities, all operating toward the realization of a utopian vision of a 'smart, connected,' HE sector based on massive volumes of student data and analysis. Beneath the utopian vision, however, is the more politicized project of market reform. Holding it together is the central issue of standardization.

Data standards

At the core of HESA's Data Futures program are negotiations about data standards. By focusing on standards in this section, the intention is to highlight how standardization functions as a political act to define how HE institutions record data, to prescribe which data they report, and thus to shape how universities may be known, perceived and potentially intervened-upon. The performance of HE market reform depends on the pre-defined standards set to enact it.

The work of standards & standardization

As Bowker and Star (2000: 13) have detailed, a standard constitutes 'any set of agreed-upon rules for the production of (textual or material) objects' which may be 'deployed in making things work together.' Embedding certain standards within a large scale infrastructure operationalizes certain rules of production and practice. The control of standards is therefore a central feature of social, economic, cultural and political life because they integrate into the infrastructures that define and participate in coordinating and organizing institutions and even whole societies (Slota & Bowker, 2017). Standards define the switches, sockets and interoperable connectors that enable an infrastructure to be assembled and operationalized, but are not just technical specifications, however. Busch (2011: 2) has noted that 'standards shape not only the physical world around us but our social lives and even our very selves. Indeed, standards are the recipes by which we create realities.' Furthermore, Star and Lampland (2009: 5) add that standards are 'increasingly linked to and integrated with one another across many organizations, nations and technical systems,' and that they 'codify, embody, or prescribe ethics and values, often with great consequences for individuals (consider standardized testing in schools, for example).' In short, then, standards are agreed rules for making things work together; defining and controlling standards is a political act; standards are stretched across and orchestrate organizations, nations and technical systems; and they are consequential for how people and things are categorized and thereby understood and treated as a result.

With the emergence of networked information and communication technologies (information infrastructures) in recent decades, new kinds of standards have been defined to organize, categorize and store the enormous quantities of data that these systems and their users produce. Data standards are an essential prerequisite to any large scale information infrastructure because they provide benchmarks for data quality and define how information are formatted and categorized in order to be stored, managed, searched for and used across many different software applications (Bowker, 2008). Moreover, as Kitchin (2014: 19) has detailed, data are recorded into forms and measures conformant with standards that have been debated, negotiated, invented and designed by

people to perform a specific task. In other words, data standards define what will be visible within a dataset and how those data may be categorized, joined-up, combined and analysed. The definition of standards in an HE data infrastructure project such as Data Futures, then, is a way of controlling which other data producing devices may be plugged in, of which data may be generated and shared, of which analyses are permitted, and of which activities and phenomena may be monitored and reported. Bowker (2008: 111-12) has noted:

> It is not only the bits and bytes that get hustled into standard form in order for the technical infrastructure to work. People's discursive and work practices get hustled into standard form as well. Working infrastructures standardize both people and machines.

As such, 'the development and maintenance of standards is a complex political and philosophical problem' since 'standards undergird our potential for action in the world, both political and scientific,' and 'make the infrastructure possible' (Bowker, 2008: 116–17).

Standardizing HE data

HESA is proposing new data standardization systems as part of its efforts to coordinate a new data infrastructure for Higher Education in the UK. As noted above, Data Futures is an operational response to the earlier HEDIIP project that specified a 'new data landscape' for HE in the UK. One of the key findings of HEDIIP was that there is an absence of commonly agreed data standards across the HE sector, with each stakeholder group—including HE providers, funders, and regulators—all acting as data providers with their own standards to meet their operational needs (KPMG, 2015). As a result, data accessible to government departments and agencies, non-governmental bodies, the media, the public and students themselves from the HE landscape was perceived to be uncoordinated and high in redundancy and duplication, with a lack of consistent data sharing between data providers and data collectors.

With the development of Data Futures, HESA has committed to the development and maintenance of a 'common data language for HE data collections,' as part of its new 'architecture for the information landscape' of HE in the UK (https://www.hesa.ac.uk/innovation/language). The standardized common data language means that data collected about student finance and applications, information provided by funders and regulators, as well as data from national leadership programs and business intelligence may be combined to produce rationalized data flows and enable enhanced data analytics and intelligence. As a result, the data standards underpinning Data Futures will offer 'common data definitions that can be used across the landscape to make reporting more efficient and make published information more comparable.'

Data Futures will also rationalize the 'standard dataset' to be collected from HE providers. HESA has collected 'student record' data since 1994, with the standard dataset described as the collective name for all the items required in the new data landscape. As part of this, Data Futures involves refining the HE student data model—a diagrammatic model of all the various items of student data required from each institution for processing. The standard dataset is also intended to enable the migration of historical data into the new model to support time series reporting. A technical specification for the data

collection platform released by HESA for potential suppliers indicates how the standard dataset would need to be both human- and machine-readable (https://www.hesa.ac.uk/files/Data-futures/Data-Futures_Overview-for-PQQ.pdf). As such, the standard dataset requires both a standardized language for human comprehension and standardized coding for software processing.

Through Data Futures, HESA has been positioned as a centralized governance body for the maintenance of these data standards and standard datasets across the UK. As the original KPMG report on the new HE data landscape report recommended:

> To the greatest extent possible, data collections are centralised in the 'transformed HESA' and other Data Collectors obtain their data from HESA, via appropriate agreements. HESA will collect the Standard Dataset in the first instance, but there is scope for non-standard data to be collected on behalf of the Data Collectors. Process changes will focus on HESA becoming the single collector of the Standard HE dataset … and other Data Collectors collecting the standard HE data from HESA. (KPMG, 2015: 11)

In this model, the 'transformed HESA' will act as a central standardizing body and an intermediary between HE institutions, which act as data providers, and all other data collecting agencies. Consequently, HE providers 'would not need to maintain multiple data relationships with multiple organisations, nor would they need to submit different sets of data, with different data definitions, at different times' (KPMG, 2015: 68).

Global quality standards

As well as producing standards, Data Futures is also regulated by standards. One document produced during the initial HEDIIP project explicitly notes that its own standards would need to be compatible with existing standards defined by the International Organization for Standardization (ISO). In particular, HESA's new standards and data language would need to refer to ISO 9001, an international standard for quality management systems, and the international security standard ISO 27001 (Youell, 2015b). Easterling (2016, 171–72) has described ISO as a global 'meta-organization,' a 'crossroads for nearly every type of organization in the world,' and as the 'beginnings of a "world state"' that 'formats the performance and calibration of many components of infrastructure … at every scale, from the microscopic to the gigantic.' Its ISO 9000 family of quality management standards represents an attempt to impose uniform management and quality assurance processes on organizations worldwide, based on existing management theories pertaining to 'the process of production, the procedures and practices of a company' and their 'social architecture,' and has catalysed the development of s global consultancy industry regarding standards compliance (Easterling, 2016: 187). Compliance with ISO 9000 standards involves an organization evaluating itself in terms of its objectives, such as customer satisfaction, and often leads to 'obsessive data gathering and metrics … to quantify or prove that deliberate objectives have been met' (Easterling, 2016: 187). As Star and Lampland (2009) have noted, standards are always nested within other standards.

The standards defining data collection in UK HE under HESA's proposed new data landscape, then, must be nested within internationally-defined standards pertaining to quality and security. In particular, driven by the political objective of putting 'students

at the centre' of a marketized HE, Data Futures is to be evaluated against a standard of quality management whereby it must evidence the achievement of that objective. Putting students at the centre is therefore framed by a uniform global standard that itself emphasizes data gathering, metrics, and the production of evidence about internal processes. HESA's compliance with ISO 9001 imposes a further standard of performance measurement on top of its objectives to standardize the collection and dissemination of student data. In other words, HESA will be required to record quality management metrics about the performance of its Data Futures metrics.

Standards are an important focus in infrastructure studies because they serve as gateways between disparate sociotechnical systems and thereby support the linking of systems into networks that might catalyse change and shape an existing social order (Slota & Bowker, 2017). The data standards developed by Data Futures are designed to interlink institutional systems into massive interoperable networks of student data collection, analysis and dissemination, in ways defined to ensure that performance towards government-prescribed market reforms can be recorded and reported in standardized, measurable and comparable metrics. These standards, in turn, are nested in global quality management standards. Standards reveal the political work of infrastructure, insofar as standards shape the practices of those working within the infrastructure and reflect the preferences, values and practices of those who defined them. Since 'standards are the rules by which we are told we should live, and the range of possibilities presented to us when we make choices' (Busch, 2011: 24), then the work of standards-building within HE data infrastructure can be understood as an attempt to govern the everyday tasks, activities and experiences of universities, and as a set of encoded rules which might regulate choices and decision-making long into the future.

Data dashboards & visualized analytics

Data standards are often invisible to those working within an infrastructure, but data visualization offers a highly visible window on to the data that is shaped by those standards. In this section, data visualization of HE is approached as a key way in which the standardized data flowing through the infrastructure are mobilized as accessible displays to shape public perception, policymakers' decisions, and HE managers' own reviews of their institutional performances. In these ways, data dashboards and visualizations function as graphical performance indicators of institutions' progress toward the accomplishment of government market reforms, and as visual prompts for certain kinds of action and intervention.

Governing through dashboards

Data visualization has proven to be a popular way for making sense of data and communicating findings, insights or meanings discovered from patterns in large datasets. Kitchin (2014: 106) has described how 'visual methods effectively reveal and communicate the structure, pattern and trends of variables and their interconnections,' thereby enabling users to navigate and query data, gain an overview of entire datasets, zoom in on items of interest, view relationships, and monitor the real-time dynamics of a phenomenon. Moreover, 'dashboards of visualized dynamic data are often on display on computer monitors in modern control rooms, summarising graphically a system in

flux for human operators, with time-series graphs and charts, or maps of unfolding events' (Kitchin, 2014: 106). Bartlett and Tkacz (2017) have described how data dashboards have migrated from the private sector to government practice and the public sector to produce a contemporary practice of 'governance by dashboard':

> Dashboards introduce new dynamics, skills, pressures, opportunities and challenges into the practice of governance. They signal a broad epistemological and organisational realignment in that they introduce new capacities to know, new criteria for what counts as good knowledge, and new ways of acting in relation to its forms of knowledge. (Bartlett & Tkacz, 2017: 8)

As tools of governance, dashboards encourage concentrated emphasis on metrics, indicators and measures, more intensified forms of monitoring and analysis, 'change the empirical basis from which decisions are made and also the criteria for what counts as a good decision,' and 'bring about a new "ambience of performance", whereby members of staff or the public become more attuned to how whatever is measured is performing' (Bartlett & Tkacz, 2017: 8).

Understood in this way, data dashboard visualizations are increasingly present as part of the 'ambience of performance' in the modern 'control rooms' of Higher Education. Dashboards enable leaders and managers to gain an accessible overview of institutional performance on many metrics, to drill down to inspect the details of specific phenomena or events, and to govern HE more effectively through visualized standards (Wolf et al., 2016). At the same time, though, how dashboards 'present this data, and how it is acted upon, in turn create new modes of behaviour, attitudes and norms within the organisations that use them' (Bartlett & Tkacz, 2017: 14).

Visualizing HE data

Data Futures exemplifies how the control room is being imported into university management offices as part of the emerging HE data infrastructure, in ways that have the potential to shape behaviours, attitudes and norms, or at the very least to attune attention to certain visualized measurements. For example, Civica Digital, the outsourced contractor appointed to deliver the Data Futures program of data collection, has the task of creating an 'improved data model and extended capabilities [which] will offer users of HESA data a regular flow of accessible information through an enhanced user interface and visualisation tools' (https://www.hesa.ac.uk/news/02-03-2017/civica-chosen-data-futures). Moreover, in its presentations detailing the new data landscape published in summer 2017, HESA indicated that its Data Platform would consist of dashboards, data tools, and data quality and standards. Its core mechanism for delivering these dashboards and visualizations is the software platform Heidi Plus, described as 'the business intelligence tool for the Higher Education sector' (https://www.hesa.ac.uk/services/heidi-plus).

Available through a subscription to HE providers and other not-for-profit organizations, Heidi Plus provides 'intuitive drag-and-drop software' enabling users 'to create interactive visualisations and dashboards to reveal the insights you need,' and to 'create competitor groups to benchmark against and choose from different metrics and filters to suit your needs.' In addition to its easy-to-use visual analytics functionality, Heidi

Plus also enables access to historic and current UK HE data which can then be used to perform time-series comparative analysis. Heidi Plus is itself the product of Tableau Server, a commercial market leader in business intelligence and analytics platforms, which markets its products as 'governed self-service analytics at scale' (https://www.tableau.com/products/server).

As previously noted, in 2015 HESA collaborated with Jisc to create a collaborative Analytics Lab. Analytics Lab is an agile data processing environment using advanced education data analytics and visualizations to create HE dashboards to enhance decision-making and strategic planning. The Analytics Lab is responsible for creating a 'Heidi Plus dashboard collection' by working with partner HE institutions to acquire 'the necessary data to develop dashboards to solve problems'. Promoted as 'a business intelligence shared service for UK education' and a 'live data processing environment,' Analytics Lab emphasizes 'cutting edge data manipulation and analysis' and enables teams 'to acquire, combine and visualise national data sets helping the higher education (HE) sector exploit the benefits of business intelligence' (https://www.business-intelli gence.ac.uk/). The task of the Analytics Lab teams has been to identify data sources and create bespoke data dashboards to allow HE providers to examine relevant data.

The dashboards and visualizations generated by Analytics Lab teams are then made available as publicly accessible 'Community Dashboards' through Heidi Plus (https://www.hesa.ac.uk/blog/12-09-2017/heidiplus-community-dashboards), described on the Analytics Lab webpage as the 'HESA national dashboard delivery service' and a tool for 'next generation education analytics':

> Analytics Labs offers higher education institutions an experimentation opportunity to refresh Heidi Plus content with insights from a wide range of data sources. Cross-institutional, analysis teams work together to solve problems using well known data sources such as the HESA collections (drawing on a Jisc/HESA data sharing agreement) linked to other educational (cross sector), demographic, employability, economic and geospatial data sets. (https://www.business-intelligence.ac.uk/analytics-labs/)

Analytics Lab teams work in a controlled analysis environment where data can be kept secure, using advanced business intelligence tools, 'to rapidly produce analyses, visualisations and dashboards for a wide variety of stakeholders to aid with decision making.' Heidi Plus is already available to subscribing institutions, with a package of supporting materials and training events, in order to enable university leaders and managers to engage in their own production of drag-and-drop visualizations in order to support decision-making. Lead users in institutions can even access personally-identifiable student and staff data. Notably, early in 2018 HESA signed an agreement with both *The Guardian* and *The Times* newspapers to use Heidi Plus to produce interactive HE dashboards of rankings and measures based on their league tables (https://www.hesa.ac.uk/news/03-01-2018/hesa-HE-league-tables). This, claimed HESA, would 'enable universities to accurately and rapidly compare and analyse competitor information at provider and subject level, changes in rank year on year,' and 'the highest climbers and the biggest "fallers."' It also noted that the dissemination and presentation of league table data help shape public opinion about different providers.

As a decision support system, the Heidi Plus dashboards and visualizations represent not just the representation of data, but a form of 'visual reasoning' known as 'visual analytics' (Kitchin, 2014: 109). Visual analytics involves humans working with algorithms to extract information and build visual models and explanations, such as through interactive visualizations that can be manipulated and used to explore and reveal patterns and connections, or that can be combined and 'mashed' with other applications for further exploration and analysis (Kitchin, 2014). Heidi Plus is an operational instantiation of, and a visible interface to, the data infrastructure of HE in the UK, a software platform supported by technical specialists, standards and visual analytics methods that is able to render in visible media progress and outcomes in relation to diverse metrics, and participate in institutional decision-making processes.

Dashboards as political indicator systems

In combination, Heidi Plus, Tableau Server, Civica Digital and Analytics Lab operationalize the practices of the 'control room' within HE institutions. Infrastructure studies tend to emphasize the hidden and invisible aspects of infrastructure, as being 'sunk into, inside of, other structures, social arrangements, and technologies' (Bowker & Star, 2000: 35). However, the data visualization and dashboard technologies developed to support Data Futures act as highly visible and legible interfaces to the new data infrastructure of UK HE. In addition, they act as bearers of visualized standards and as graphic portals to standard datasets which are able to apply agreed-upon rules and definitions to the analysis of the social realities of the things and people that inhabit universities. Understood in this way, the data dashboards and visualizations developed as part of Data Futures act as control room technologies for monitoring institutional progress toward key performance indicators as defined by the Higher Education and Research Act (HERA). As political indicator systems, the Data Futures dashboards both enable institution to experiment with data and its presentation, while also allowing institutional data to be visually represented for policymakers, publics and regulators, and thereby to shape decision-making and other forms of action and intervention.

An identified problem with visualizations and dashboards is that they reduce and simplify complex phenomena for easy interpretation:

> Data presented on a dashboard is rarely as straightforward as it appears. Dashboards condense data for easy digestion, which can obscure a user's knowledge of how trustworthy or accurate that data is. By presenting often very complex, messy and varied data in simplified forms for consumption via a dashboard, sometimes subtle changes take place in how that data is understood. (Bartlett & Tkacz, 2017: 5)

These visualized forms transform complex matters, information and data into simplified visual arrays and representations. As Latour (1986) has argued, visualisations act as material techniques of thought. They can be moved around, copied, reshuffled, recombined, superimposed and reproduced in other places. The power of a visualisation, graphic, image, or diagram, is to stabilize ideas, problems, concepts, explanations and arguments in one place so as to influence the way people think about them. Subtle

cues in colouring, the design of charts and graphs, and 'other visual cues all guide the user's attention to preferred interpretations of the data on display' and thereby shape 'the priorities of its user' (Bartlett & Tkacz, 2017: 15).

Dashboards also bring new forms of expertise and authority into decision-making processes. The creation of dashboards and the visualizations they present is the expert accomplishment of graphics designers, data analysts, algorithm specialists, whose own practices are therefore entangled in those of the phenomena they are tasked to make visible and interpretable (Rose, Degen, & Melhuish, 2014). In this sense, as data dashboards and visualizations proliferate across HE, new kinds of privileged positions are being attained by technical experts who are both able to manipulate the data and also narrate what those data are 'saying' by translating complex numerical calculations into stable and easy-to-interpret diagrammatic form.

The turn to governance by dashboard in HE exemplifies a larger shift of governance in relation to big data. Davies (2017: 2) argues that traditional forms of expertise, authority and judgment have experienced diminishing levels of public trust in recent years, with 'experts' and 'elites' critiqued over their lack of 'objective judgment over the "facts" of what is taking place.' Instead, computational systems which can process big data, and those who manage them and report what the numbers mean, are gaining increasing public status and authority. As a consequence, expert power now increasingly resides in a combination of nonhuman, real-time feedback technologies and the human intermediaries who can translate and narrate the flow of data to make it intelligible for 'public audiences' and 'political agents and states' (Davies, 2017: 17). In other words, 'the rise of big data privileges those capable of mediating between mathematical analytics and empirical narratives about what is being represented' (Davies, 2017: 18). As such, Data Futures requires new forms of expertise in the handling of political indicator systems by which HE institutions may be judged and held accountable in terms of their market performance.

Within HE, new kinds of real-time feedback technologies such as business intelligence applications and visualization tools such as Heidi Plus are becoming powerful nonhuman intermediaries which, alongside the human actors who can manipulate and interpret them, are attaining a privileged position to define how HE data are understood, made actionable, and acted upon. Underpinned by standards defined in accordance with specific political, social and economic priorities, and projected into the world via dashboards and visualizations, the emerging data infrastructure of HE in the UK has the potential to reshape how universities are governed, and thereby to reshape organizational and individual behaviours within the new ambience of measurement and performance.

Discussion: The 'extrastatecraft' of infrastructure

The Data Futures program of the UK Higher Education Statistics Agency is a major national infrastructure building project, an attempt to construct a material substrate for data collection, processing and sharing that might support a utopian vision of big data-driven, digitized and smarter universities while also building the architecture for market reform of the HE sector. A huge global industry in the production of big data technologies for HE has emerged—from organizational and business intelligence to learning analytics—which can be plugged into the sockets of smarter universities.

Emerging technologies of machine learning, adaptive learning platforms and even artificial intelligence are all being promoted as future technologies that will enhance Higher Education, though as with all big data applications they will require 'agile' infrastructure to plug into (Adams Becker et al., 2017). Around the world, infrastructural utopias of smarter universities are being developed, often promoted through a mixture of political activity, think tank reports, commercial advocacy and global consultancy. This paper has examined Data Futures as an infrastructural project currently in-the-making that is establishing an agile architecture for future big data-driven technological add-ons, and in so doing establishing the infrastructure for completion of the market reform of HE demanded by the 2017 Higher Education and Research Act. There remains an important need for comparative studies analysing how new dynamic, data-processing infrastructure is being sunk into, inside of, other national Higher Education systems and arrangements.

In the US, for example, the Institute for Higher Education Policy (IHEP) is leading a Postsecondary Data Collaborative known as PostsecData. IHEP itself is a Washington, DC-based 'nonpartisan, nonprofit organization committed to promoting access to and success in higher education for all students,' which 'develops innovative policy- and practice-oriented research to guide policymakers and education leaders who develop high-impact policies that will address our nation's most pressing education challenges' (http://www.ihep.org/about-ihep). Supported with funding from the Bill and Melinda Gates Foundation and first convened in 2015, the PostsecData collaborative is seeking to develop 'robust and impactful postsecondary education data policies' and to 'inform how policy leaders make critical decisions about what data to collect, how to collect it, who should have access to it, how to define metrics, and how to present data to the public' (http://www.ihep.org/postsecdata/about).

As with Data Futures in the UK, PostsecData also proceeds from the view that existing infrastructural arrangements are inadequate for emerging needs:

> the existing national postsecondary data infrastructure … is burdensome,
> uncoordinated, and increasingly at risk of slipping into obsolescence. To move forward
> as a nation, we must take the opportunity now to create an agile and effective national
> postsecondary data ecosystem whose individual components communicate with and
> build upon each other to enable all stakeholders in the enterprise to focus on what
> really matters: student success. (Cubarrubio & Perry, 2016: 11)

In 2017 Postsecdata proposed a new federal Student-Level Data Network (SLDN), to be hosted by a central statistics agency with governance responsibilities for data collections and standards, which would 'require significant federal action and interagency collaboration' (Roberson et al., 2017: 9). Like Data Futures, the PostsecData project to produce a federal SLDN is not the ambition or accomplishment of one single organization, but a more or less organized coalition of partners and relationships that crisscross the public and commercial sectors, and includes think tanks, non-profit institutions, academic institutions, consultancy groups, lobbying alliances and wealthy philanthropic foundations. At the core of IHEP's efforts is a series of proposals and targeted recommendations 'envisioning the national postsecondary infrastructure in the 21st century' (http://www.ihep.org/postsecdata/mapping-data-landscape/national-postsecondary-data-infrastructure). It emphasizes an

infrastructure with student needs at its centre, and raises a series of challenges to overcome. These include data governance (who owns the data and the systems that collect, analyze, and disseminate them?); data use (how can the ability for data to be reported to the public and the community of practice be improved, and how can institutions' use of data for improvement and accountability be promoted?); and standards and interoperability (how can different system owners work together to ensure interoperability and harmonize the use of different data standards and definitions by different data collection entities?). Although there are clearly contrasting local differences between the UK and US, both Data Futures and PostsecData are premised on a shared utopian vision of digital data-driven HE in the twenty-first century—a vision both projects are seeking to operationalize through new data infrastructure models.

The emergence of proposed new data infrastructures for Higher Education that make use of big data, new standards, and data analytics, dashboards and visualizations is part of a shifting infrastructural landscape where the work of humans and nonhuman technologies is increasingly hybridized into assemblages, with infrastructure understood as 'integrally a social, organizational, and physical phenomenon' (Slota & Bowker, 2017: 537). Within such assemblages, a highly diverse web of people, organizations, technologies, standards, political and economic priorities, discourses and practices press upon and interpenetrate one another. Although Higher Education has historically been part of the publicly funded state, with increased marketization and the introduction of new commercial providers, HE is now extended beyond the state to become part of the global movement of consulting firms and commercial technology companies, and has fused with the managerial discourse of marketization from the private sector (Selwyn, 2014).

The emerging data infrastructure of HE in the UK is therefore not just the product of national 'statecraft,' but of what Easterling (2016: 15) terms 'extrastatecraft':

> Contemporary infrastructure … is the secret weapon of the most powerful people in the world precisely because it orchestrates activities…. Some of the most radical changes to the globalizing world are being written … [in] infrastructural technologies—often because market promotions or prevailing political ideologies lubricate their movement.

The extrastatecraft of infrastructural technologies is especially pronounced with regard to technologies of data collection. While state agencies and authorities have historically maintained control over national data, now the 'monopoly of the state over data production, collection, and even interception is increasingly challenged … by corporations, agencies, authorities, and organizations that are producing myriad data' (Ruppert, Isin, & Bigo, 2017: 3–4). Data has therefore become 'an object whose production interests those who exercise power' because 'there has never been a state, monarchy, kingdom, empire, government, or corporation in history that has had command over such granular, immediate, varied, and detailed data about subjects and objects that concern them' (Ruppert et al., 2017: 3). Control over data infrastructure, then, is a mode of data politics since it confers great powers for infrastructure owners to be able to know, measure and fit into standard categories those subjects and processes that are integrated into the infrastructure.

Companies such as Civica Digital and Tableau Server have thus become essential partners for state departments of government and their agencies, such as the UK

Higher Education Statistics Agency, the Department of Business, Innovation and Skills, and the Office of Students. These extra-state actors strengthen the optical powers of the state to see into and inspect its own institutions, and to enact political legislation such as the Higher Education and Research Act. In particular, Civica Digital and the Heidi Plus software package provide the state with more fine-grained techniques by which to render HE institutions and the people who occupy them in statistical, visualized, comparable, and evaluative terms, all benchmarked to particular standards. These software tools are in an important sense parts of the extrastatecraft of contemporary Higher Education infrastructure, acting as sociotechnical relays for the data-driven practices required by smarter universities in a marketized sector. They are integral to HESA's position as a centre of calculation that can collect and 'warehouse' data from large numbers of institutions, and then process and distribute those data to other sites in ways that might produce conviction about what those data are saying, and thereby shape decision-making, political deliberation or policymaking processes.

Conclusion

The Data Futures program is building the data architecture, defining the language and standards, establishing the relevant practices of real-time analytics and visualization, and providing the sockets to plug in other emerging big data technologies, that, as an infrastructural substrate, will sink into and orchestrate UK HE institutions as the Higher Education and Research Act is enacted in future years. It is both a mundane and utopian project. HESA's efforts to build a new HE data infrastructure are motivated by contemporary political contests over the governance of the university sector, and it is expected to make institutional data more accessible to potential fee-paying student applicants as a way of developing a competitive marketplace of providers. It is also designed to speed up the collection of teaching quality data and enable UK universities to be evaluated and ranked. Powerful and politically-connected figures such as Michael Barber, chair of the new Office for Students, are driving forward a reformatory agenda that puts market competition and metrics at the core of HE, with Data Futures as its hidden architecture for data collection and measurement. To some extent, Data Futures is a centralized data surveillance project, casting a grid of statistical visibility across the HE sector in order to enable governmental centres of calculation to calculate about institutions, and incite institutions to calculate about themselves according to the same statistical standards. Together, these infrastructural elements are paving the way for 'human-algorithmic decision-making' and surveillance in HE (Prinsloo, 2017). But Data Futures is also, as its chief executive has claimed, part of a utopian vision of 'a digital HE sector, with data-driven universities operating within a smart, connected environment,' where institutions can self-monitor, policymakers can analyse sectoral data, and users can measure their progress in real-time (Clark, 2016). In this sense, Data Futures supports the government's 2017 Industrial Strategy, which emphasizes HE institutions' role in innovation and the production of 'digital skills' for the future. It exemplifies efforts underway around the world to design and build smarter, data-driven and market-competitive digital universities for the future (Lane & Finsel, 2014).

Underpinning the ideal of the marketized smarter university is a complex mosaic of people, technologies, standards and policies, all of which are being brought into

alignment as the social, political, technical and material substrate to the utopia of a big data-driven, marketized Higher Education sector. Similarly driven by utopian ideals of infrastructural reform, HE systems in other parts of the globe are beginning to shift toward agile data infrastructure arrangements which will permit new sources and practices of big data to be integrated into organizational and pedagogic processes as well as reformatory programs. Business intelligence applications, education analytics, adaptive learning platforms and other 'smart learning tools,' as well as data dashboards and visualizations used by HE leaders and policymakers to support decision-making processes, are set to be plugged into the architecture of the university, in ways that will impose new modes of quantification and standardization while also bringing new actors and priorities from across the public and private sectors into contemporary HE. As Selwyn (2014: 49) notes, 'these systems should be seen as a tangible embodiment of the recasting of universities over the past two decades along more business-orientated, centralized "data-driven" lines.' Data Futures exemplifies how Higher Education is coming to be governed and regulated through the simultaneous statecraft of market reform and the extrastatecraft of sociotechnical infrastructure. Specific software programs, standards and private sector organizations that facilitate live data and real-time metrics are fusing with political reforms in the shaping of a marketized sector of smarter universities.

Author information
Ben Williamson is a lecturer at the University of Stirling, UK. His research focuses on digital technology and education policy, and his latest book is *Big Data in Education: The digital future of learning, policy and practice*.

Competing interests
The author declare that he/she has no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
Adams Becker, S., Cummins, M., Davis, A., Freeman, A., Hall Giesinger, C., & Ananthanarayanan, V. (2017). *NMC horizon report: 2017 higher education edition*. Austin, Texas: The New Media Consortium.
Bartlett, J., & Tkacz, N. (2017). *Governance by dashboard: A policy paper*. London: Demos.
BIS (Department of Business, Innovation & Skills). (2011). *Students at the Heart of the System*. Gov.uk, June 2011: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/31384/11-944-higher-education-students-at-heart-of-system.pdf
BIS (Department of Business, Innovation & Skills). (2016a). *Success as a Knowledge Economy: Teaching Excellence, Social Mobility and Student Choice*. Gov.uk, May 2016: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/523546/bis-16-265-success-as-a-knowledge-economy-web.pdf
BIS (Department of Business, Innovation & Skills). (2016b). Case for the creation of the Office for Students. Gov.uk, June 2016: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/527757/bis-16-292-ofs-case-for-creation.pdf
Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. London: MIT Press.
Bowker, G. C. (2008). *Memory practices in the sciences*. London: MIT Press.

Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.

Burnett, K. (2017). By creating a market for universities, the government has snookered itself. *The Guardian*, 7 September: https://www.theguardian.com/science/political-science/2017/sep/07/by-creating-a-market-for-universities-the-government-has-snookered-itself

Burrows, R. (2012). Living with the h-index: Metric assemblages in the contemporary academy. *The Sociological Review*, *60*(2), 355–372.

Busch, L. (2011). *Standards: Recipes for reality*. London: MIT Press.

Carrigan, M. (2017). The coming big data revolution within higher education. Markcarrigan.net, 17 October: https://markcarrigan.net/2017/10/17/the-coming-big-data-revolution-within-higher-education

Clark, P. (2015). The Green Paper needs big data. *Times Higher Education*, 18 November: https://www.timeshighereducation.com/blog/green-paper-needs-big-data

Clark, P. (2016). Higher education and the data revolution, 25 April: http://www.citethisforme.com/blog/2016/04/25/higher-education-and-the-data-revolution

Colvin, C., Rogers, T., Wade, A., Dawson, S., Gasevic, D., Buckingham Shum, S., … Fisher, J. (2016). *Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement*. Canberra, ACT: Australian Government Office for Learning and Teaching http://www.olt.gov.au/project-student-retention-and-learning-analytics-snapshot-currentaustralian-practices-and-framework.

Cubarrubio, A. & Perry, P. (2016). Creating a thriving postsecondary data ecosystem. IHEP PostsecData resources, May 2016: http://www.ihep.org/sites/default/files/uploads/postsecdata/docs/resources/postsecondary_education_data_ecosystem.pdf

Davies, W. (2017). Elite Power under Advanced Neoliberalism. *Theory, Culture & Society*: https://doi.org/10.1177/0263276417715072

Deloitte. (2013). *Redesigning the higher education data and information landscape*. Strand 1 project report, March 2013: https://www.hediip.ac.uk/wp-content/uploads/RPG_HEDIIP_proposal_2013-03.pdf

Dillenbourg, P. (2016). The evolution of research on digital education. *International Journal of Artificial Intelligence in Education*, *26*(2), 544–560.

Dourish, P., & Bell, G. (2011). *Divining a digital future: Mess and mythology in ubiquitous computing*. London: MIT Press.

Easterling, K. (2016). *Extrastatecraft: The power of infrastructure space*. London: Verso.

Education Design Lab. (2014). The edtech revolution is about to become the learner revolution. Education Design Lab, April 2014: http://eddesignlab.org/wp-content/uploads/2014/04/LearnerRevolution_EducationDesignLab.pdf

Edwards, P. N., Jackson, S. J., Chalmers, M. K., Bowker, G. C., Borgman, C. L., Ribes, D., … Calvert, S. (2013). *Knowledge infrastructures: Intellectual frameworks and research challenges*. Ann Arbor, MI: Deep Blue.

Guster, D., & Brown, C. G. (2012). The application of business intelligence to higher education: Technical and managerial perspectives. *Journal of Information Technology Management*, *23*(2), 41–62 http://jitm.ubalt.edu/XXIII-2/article4.pdf.

Guy, M. (2017). What does the Data Matters Conference offer data experts? HESA Blog, 27 September: https://www.hesa.ac.uk/blog/27-09-2017/data-matters-conference

HESA. (2016). Corporate strategy 2016–21. Cheltenham: HESA, https://www.hesa.ac.uk/files/HESA_Corporate_Strategy_2016-2021.pdf

HESA. (2017). Data Futures. HESA, August: https://www.hesa.ac.uk/files/Data_Futures_overview_Aug17_final_0.pptx

HM Government (2017). *Industrial strategy: Building a Britain fit for the future*. London: Crown Copyright.

Hoel, T. (2014). Standards and Learning Analytics – current activities 2014. LACE, 15 December: http://www.laceproject.eu/blog/standards-learning-analytics-current-activity-2014/

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. London: Sage.

Kitchin, R. & Lauriault, T. (2014). Towards critical data studies: charting and unpacking data assemblages and their work. *The Programmable City Working Paper 2*: http://ssrn.com/abstract=2474112

Knox, J. (2017). Data power in education: Exploring critical awareness with the 'Learning Analytics Report Card'. *Television & New Media*: https://doi.org/10.1177/1527476417690029

Komljenovic, J. & Robertson, S.L. (2016) The dynamics of 'market-making' in higher education. Journal of Education Policy, 31(5), 622–636.

KPMG. (2015). The blueprint for a new HE data landscape: Final report. HEDIIP: https://www.hediip.ac.uk/wp-content/uploads/HEDIIP_Data_Landscape_Report.pdf

Lane, J. E., & Finsel, B. A. (2014). Fostering smarter colleges and universities: Data, big data and analytics. In J. E. Lane (Ed.), *Building a smarter university: Big data, innovation and analytics*. Albany: State University of New York Press.

Larkin, B. (2013). The politics and poetics of infrastructure. *Annual Review of Anthropology*, *42*, 327–343.

Latour, B. (1986). Visualization and cognition: Thinking with eyes and hands. *Knowledge and. Society*, *6*, 1–40.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford: Oxford University Press.

Leach, M. (2017). Higher Education Power List 2017. *WonkHE*, 20 September: http://wonkhe.com/2017-higher-education-powerlist/

Losh, L. (2014). *The war on learning: Gaining ground in the digital university*. London: MIT Press.

Lupton, D. (2015). *Digital Sociology*. London: Routledge.

Mayer-Schonberger, V., & Cukier, K. (2014). *Learning from big data: The future of education*. New York: Houghton Mifflin Harcourt.

Perrotta, C., & Williamson, B. (2016). The social life of learning analytics: Cluster analysis and the performance of algorithmic education. *Learning, Media and Technology* https://doi.org/10.1080/17439884.2016.1182927.

Plantin, J.-C., Lagoze, C., Edwards, P.N. & Sandvig, C. (2016). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*: http://journals.sagepub.com/doi/full/10.1177/1461444816661553

Policy Connect (2016). *From bricks to clicks: The potential of data and analytics in higher education*. London: Policy Connect http://www.policyconnect.org.uk/hec/sites/site_hec/files/report/419/fieldreportdownload/frombrickstoclicks-hecreportforweb.pdf.

Prinsloo, P. (2017). Fleeing from Frankenstein's monster and meeting Kafka on the way: Algorithmic decision-making in higher education. *E-Learning and Digital Media*, *14*(3), 138–163.

Ridley, D. (2017). Editorial: The myth of the market. *HE Marketisation*, (10 December) https://hemarketisation.wordpress.com/2017/12/10/editorial-the-myth-of-the-market/.

Roberson, A. J., Rorison, J., & Voight, M. (2017). *A blueprint for better information: Recommendations for a federal postsecondary student-level data network*. Washington, DC: Institute for Higher Education Policy http://www.ihep. org/sites/default/files/uploads/docs/pubs/a_blueprint_for_better_information_ihep_policy_brief_0.pdf.

Rose, G., Degen, M., & Melhuish, C. (2014). Networks, interfaces and computer-generated images: Learning from digital visualizations of urban regeneration projects. *Environment and Planning D: Society and Space*, *32*, 386–403.

Ruppert, E., Isin, E. & Bigo, D. (2017). Data politics. *Big Data & Society*, July–December: 1-7.

Say, M. (2017). HESA appoints Civica digital for data futures. *UKAuthority, data4good*, (3 March) http://www.ukauthority. com/data4good/entry/6929/hesa-appoints-civica-digital-for-data-futures.

Selwyn, N. (2014). *Digital technology and the contemporary university: Degrees of digitization*. London: Routledge.

Slota, S. C., & Bowker, G. C. (2017). How infrastructures matter. In U. Felt, R. Fouche, C. A. Miller, & L. Smith-Doerr (Eds.), *The handbook of science and technology studies*, (4th ed., pp. 529–554). London: MIT Press.

SoLAR SHEILA. (2016). Using data wisely for education futures. About SHEILA Project: http://sheilaproject.eu/about-sheila/

Star, S. L., & Lampland, M. (2009). Reckoning with standards. In M. Lampland, & S. L. Star (Eds.), *Standards and their stories: How quantifying, classifying and formalizing practices shape everyday life*. London: Cornell University Press.

Westminster Higher Education Forum. (2017). Technology in Higher Education - data analytics, the learning experience and next steps for MOOCs. Westminster Forum Projects: http://www.westminsterforumprojects.co.uk/conference/ tech-in-HE-17

Wilson, A., Watson, C., Thompson, T. L., Drew, V., & Doyle, S. (2017). Learning analytics: Challenges and limitations. *Teaching in Higher Education*, *22*(8), 991–1007.

Wolf, M., Taimurty, M., Patel, M. and Meteyer, J. (2016). The Dean's information challenge: from data to dashboard. *EduCause Review*, 28 November: http://er.educause.edu/articles/2016/11/the-deans-information-challenge-from- data-to-dashboard

Youell, A. (2015a). Redesigning the information landscape. *WonkHE*, 16 March: http://wonkhe.com/blogs/redesigning- the-information-landscape/

Youell, A. (2015b). HEDIIP: Redesigning the information landscape. UCISA seminar, January: https://www.ucisa.ac.uk/-/ media/groups/cisg/events/2015/ucas_hesaetc/ay_ucisa_seminar_2015-01-15.ashx?la=en

Youell, A. (2017). A brief history of higher education data. *WonkHE*, 28 June: http://wonkhe.com/blogs/a-brief-history-of- higher-education-data/