

New shadow professionals and infrastructures around the datafied school: Topological thinking as an analytical device

European Educational Research Journal

1–15

© The Author(s) 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14749041211007496

journals.sagepub.com/home/eer**Steven Lewis** 

Australian Catholic University, Australia

Sigrid Hartong

Helmut Schmidt University Hamburg, Germany

Abstract

Drawing upon the growing datafication of contemporary schooling, our purpose in this article is to use topological thinking as an analytical device to better understand the professionals and practices within emergent data infrastructures. We address this by attending to an influential national (and subnational) data infrastructure of school monitoring in the United States, managed by the federal Department of Education, known as *EDFacts*. Informed by policy documents relating to *EDFacts*, as well as by various related software platforms and portals, we explore the *whom* and *how* of datafication, and expose the increasing presence and influence of otherwise ‘hidden’ technology mediators, or ‘shadow professionals’. In particular, we argue that the increasing dependency of *EDFacts* on data has necessitated the introduction of new professional roles associated with optimising the flow of data, and thus stabilising and normalising the topological space of the infrastructure. We conclude by suggesting that *EDFacts* encourages teaching professionals and shadow professionals alike to engage in *acts of data submission*; that is, providing data to *EDFacts* and, at the same time, positioning themselves as wholly responsive to the infrastructure and its datafied renderings of schooling.

Keywords

Data, datafication, topology, infrastructure, *EDFacts*, policy, ‘shadow professionals’

Corresponding author:

Steven Lewis, Senior Research Fellow and ARC DECRA Fellow, Institute for Learning Sciences & Teacher Education, Australian Catholic University, 229 Elizabeth Street, Brisbane, QLD, 4000, Australia.

Email: steven.lewis@acu.edu.au

Introduction

Responding to a rising body of critical research on the ongoing datafication of schooling (Hartong, 2018; Landri, 2018; Lewis and Holloway, 2019; Takayama and Lingard, 2019; Williamson, 2017), this article seeks to better understand the new infrastructures, people and practices emerging around the increasingly ‘datafied’ monitoring of schools and schooling systems. We use the term *data infrastructures* here to describe socio-technical assemblages (i.e. people and computing hardware/software) that laboriously translate things into data, with these data in turn providing the means to fabricate supposedly objective knowledge for the purpose of governance. Taken collectively, these processes of data *infrastructuring* (Piattoeva and Saari, 2019) are especially significant for their ability to enable new types of professional knowledge and, relatedly, *professionals* to occur. This concerns not only *who* is allowed to produce and, particularly, decide on the quality of monitoring data, as well as *how* they may do so, but also how these (inter)actions produce and increasingly normalise new relational, or *topological*, spaces of data measurement and governance (Gulson and Sellar, 2019; Hartong, 2018; Lewis et al., 2016). Drawing on our previous work that employs topological thinking to understand contemporary schooling spaces (for instance, see Hartong, 2018; Hartong and Piattoeva, 2019; Lewis et al., 2016; Lewis, 2020a, 2020b), our purpose in this article is to better understand the role of emergent data professionals and practices within data infrastructures, and their roles ensuring the availability, quality and flow of data. Rather than topology being deployed only as a theoretical lens, as is often the case in education policy sociology, we seek to demonstrate how topology can be developed both analytically and methodologically.

We address this by attending to an influential national (and subnational) data infrastructure of school monitoring in the United States managed by the federal Department of Education and known as *EDFacts*. Initiated in 2007, EDFacts has sought to position ‘performance data *at the centre of policy, management, and budget decisions* for all K-12 [i.e. primary and secondary school] education programs’ (U.S. Department of Education, 2020: np; emphasis added), providing the means to centrally collect, analyse and report on data supplied by schools, schooling districts and K-12 state (i.e. subnational) education agencies (SEAs). EDFacts has helped to consolidate federal regulations on state reporting requirements when receiving money from federal education programmes (e.g. through Title 1¹), including taking punitive action (e.g. withholding funding) against states that fail to report data in the prescribed manner. Moreover, it has transformed an immense amount of paperwork into a standardised, and mandatory, digital data submission system (U.S. Department of Education, 2018: 7). Since then, the EDFacts initiative has consistently expanded, gradually integrating increasing quantities of data, while also becoming increasingly streamlined through tremendous investments in technical and procedural standardisation. Of especial interest here, the increasing dependency of EDFacts on data flows (i.e. maintaining on-time inputs and outputs in a particular form) has necessitated the introduction of multiple professional roles (e.g. so-called ‘data stewards’, ‘data commissioners’ and ‘error managers’). As we will show, these individuals serve the data infrastructure by continuously controlling and setting (new) standards, as well as developing highly specialised data-centric languages for collecting, regulating and reporting on data.

Our purpose focusing on these emergent actors is to explore the *whom* and *how* of datafication, and expose the increasing presence and influence of otherwise ‘hidden’ technology mediators (see Hartong, 2016) to enable a deeper understanding of data infrastructures and their complexity. As our analysis shows, the growing focus on data infrastructuring in EDFacts has caused a significant shift in priorities, in which optimising the flow of data, and thus stabilising and normalising the topological EDFacts space, has become the primary concern of those who seek to govern the various US schooling systems (i.e. federal, state, district), rather than emphasising the actual

educational practices and outcomes of teachers or students notionally represented by these very data. This focus on a data-based rendering of schooling reflects the increasing imbrication of the social and technical characters of digital infrastructures, where standardising languages and expectations around data help to bridge the distance between the technical and the human to make data legible, usable and influential. Our purpose is to provide empirical insights into an increasingly powerful education monitoring infrastructure and, at the same time, show how topologically informed thinking can help us make empirical cases of datafication and digitalisation traceable and analysable.

We begin our article by first outlining how topological thinking is useful for understanding the discourses and activities made possible by data infrastructures, and how these have led to new topologies in education governance. We then turn our attention to an empirical analysis of EDFacts, including the historical context of the programme, the various components used for data collection and, central to our argument here, the rise of so-called ‘shadow professionals’ whose roles are to support the infrastructure by maintaining flows of ‘quality’ data. Finally, we conclude the article by suggesting that EDFacts reflects a logic and acts of *data submission*. This is both in the sense that data are required to be constantly fed into the infrastructure, while also, at the same time, we can see the relative diminution of professional agency and expertise in the face of datafied (notional) certainty.

Theoretical framework: A topological view on data infrastructures

Datafication and data infrastructures

The increasing enfolding of data within spaces, relations and processes of governance has become prevalent in contemporary society, and schooling has by no means been immune to such developments (see Lewis, 2020a). Lawn (2013: 8) notes that ‘the creation and flow of data has become a powerful governing tool in education’, while Gulson and Sellar (2019: 350) argue that datafication has led to ‘changing relations of power and space in education’. This encompasses the logic that education, teaching and learning can and should be measured, and that policy and practice decisions should be made on the basis of these data. This, however, requires efficient and effective organisation of data as data infrastructures, which refers to the various objects and subjects assembled around data collection, storage, visualisation, and mediation between otherwise disparate, diverse and disconnected actors and spaces (e.g. schools, districts, states, systems, countries).

In this contribution, it is that process of *infrastructuring* as a key rationality and technology of data-based governing in education (Hartong and Piattoeva, 2019) which we seek to analyse more closely. We argue that the underpinning logic of a data infrastructure is its emphasis upon data *flows*. Put differently, the ability of data to flow in multiple directions (from places of creation to places of use, from producers to consumers, on to subsequent consumers – who will in turn produce their own data; etc.) becomes the definitional purpose of the infrastructure. This also means, however, that, both materially and figuratively, the data infrastructure is *coterminous* with the systems, spaces and people that it represents and enables. If the data are unable to be collected from or sent to a certain person or place, then these people and places, by definition, are excluded from the infrastructure.

Furthermore, it is this ‘flow-ability’ that allows data to inform practices and decisions throughout the infrastructure, which can then generate more data for these (data-informed) changes to be evaluated and fed back to the infrastructure. The infrastructure is thus wherever and whomever and whatever (1) can be touched by data; *and then* (2) generate data themselves by the impact/use of those data; *and then* (3) input those data (and their own self-generating data) back into the system.

The infrastructure, and the ability of data to flow within it, inscribe the space as perceptible and governable (see also Hartong, 2018), and hence render it ‘real’. Just the ability to generate or use data by oneself is insufficient; it is arguably akin to shouting into a vacuum without a medium to carry the sound. Rather, the *joining up* of the people, things and spaces through, by and as data is the prime consideration.

This is a key point of the infrastructure, insofar as it embodies more than just a thing *within* a pre-existing political space. Reflecting Kitchin’s (2014) argument that data assemblages are ontogenetic, we see a given data infrastructure and its associated flow-ability to constitute that very space (i.e. ontologically), as well as to define the types of knowledge, expertise and discourses that are valued and authorised within that space (i.e. epistemically). Even though the space in question overlaps with the spaces data are derived from or fed back into, it is still new because it comprises new relations and interactions that would be unavailable *without the infrastructure* (see also Gulson and Sellar, 2019: 359). That being said, it is important not to conflate the creation of new data-driven governing spaces with the wholesale replacement of schooling systems and departments of education. Rather, the emergence of data infrastructures in education require researchers to acknowledge ‘the possibility of schools and systems as multiple spaces in which existing relations are transformed by and within data infrastructures’ (Gulson and Sellar, 2019: 362).

Consequently, it is necessary to consider *how* the various aspects of schooling performance, student learning and teachers’ work are captured as data from a myriad of dispersed sites, and then translated into a format legible to the infrastructure. Ratner and Ruppert (2019: 2) have described these efforts as ‘aesthetic practices’, which provide the means of ‘purging data of inconsistencies, differences and uncertainties’, not only to ensure they are ‘clean’ (i.e. accurate) but also to ensure they are in a form amenable to their ultimate use (e.g. informing decisions on teacher performance). We can see in such practices of translation an attempt at standardisation and sense-making, whereby putatively ‘messy’ data are corrected to ensure that the data infrastructure, including the people who engage with the infrastructure, read and make sense of the data in the same way. It is arguably these ‘little analytical devices’ (Amoore and Piotukh, 2015) that help to make big data perceptible and usable for purposes of knowing and governing social domains such as education. While these little analytical devices might well be machine-reading processes, algorithmic analysis and data mining, they equally include the people – for example, professionals – who undertake and uphold these activities. Attending to the socio-technical sense of the infrastructure, it is perhaps unnecessary (or even impossible) to distinguish between human and machine agents, so long as the data they make available are in a form that can ensure ‘something of interest is brought to attention for action’ (Amoore and Piotukh, 2015: 344).

Viewing data infrastructures through a topological lens

Despite the increasing ubiquity of education data infrastructures, the analytical and methodological tools available to policy sociologists have not always proved suitable to the task. For instance, apprehending and analysing the presence and role of data flows, and the datafied professionals and activities that facilitate such flows, is made more difficult by having to refer to the in-between relations that constitute digital infrastructures, rather than a specific site or place. It is this focus on relationality that suggests topological thinking may offer a productive way forward. Informed by the recent deployment of relational, or topological, thinking in the social sciences, Lury et al. (2012) argue that contemporary social life is increasingly marked by the growth of practices (the *becoming topological of culture*) that provide fertile ground for deploying topological thinking as an analytical tool, both in everyday life and in social and cultural theory (*cultural topology*). This ‘becoming topological’ is evident through the proliferation of dynamic ‘practices of ordering,

modelling, networking and mapping that co-constitute culture, technology and science' (Lury et al., 2012: 5), producing a spatiotemporal continuity and relationality across social domains that extends beyond the fixed rigidity of Euclidean spaces or territorial polities.

Technology and communication media, along with greatly enhanced computational capacities, play a central facilitative role here, helping to create new continuities and relations through which various *flows* of ideas, practices – or data, in this case – can diffuse within and across these emergent topological spaces:

... the becoming topological of culture is necessarily a concern with how the computational transformation of technical machines and media into systems of organisation, storage, transmission and control of information has led to a new form of culture defined by *flows of data*, and by the rules, procedures, [and] constraints through which they are ordered. (Lury et al., 2013: 2; emphasis added)

The ability of technology to order these data flows, such as locating individuals in 'joined-up' government databases – or infrastructures – of digital transaction data, is undoubtedly a key enabling feature of topological dynamics (e.g. Grommé and Ruppert, 2019; Hartong, 2018; Lewis, 2020a; Prince, 2016; Ruppert, 2012). While there has been much written about governance by numbers and data in policy sociology literature (see, for instance, Grek, 2009; Lingard, 2011; Ozga, 2016; Williamson, 2017), a topological lens specifically emphasises how governing spaces are constituted through these datafied relations, rather than merely occurring in preformed territories or scalar spaces. In short, if more things can increasingly be rendered as data, topology helps direct our attention to the diversity, complexity and ongoing evolution of the spaces and relations used to know and govern these things.

At the same time, however, these spaces and relations can only be durable – and governable, in the sense that they can enable ongoing effects – by the other character of topology; namely, that these constitutive relations and spaces can endure *in spite of* continual change or renegotiation. As Martin and Secor (2014: 431) usefully remind us, 'topology directs us to consider relationality itself and to question how relations are formed and then endure despite conditions of continual change'. In this way, topology encourages education policy research to consider not only *what* relations are present and actively constructing spaces amenable to being governed, but also *how* these relations are formed and maintained. With regard to data infrastructures, topological thinking directs our analytical and methodological attention not only to the *relations* of the infrastructure and to the governing logics mobilised and enabled by the *joining-up* of the infrastructure, but equally to the mechanisms (social, cultural, discursive, technical etc.) used to *sustain* these relations, which together creates the actual topology of the infrastructure. In other words, understanding a data infrastructure's topology means to consider how the ongoing *flow-ability* of data (i.e. change) and ongoing processes of *re/de/bordering* (i.e. maintenance; see also Robertson, 2011; Scheel, 2020; Mezzadra and Neilson, 2012; Billé, 2018; van de Oudeweetering and Decuyper, 2019) work together.

In the following section, we seek to illuminate how such a topological approach can be used not only as a theoretical set of metaphors, but rather as an *empirical* device, which necessitates a shift in both *methodological* and *analytical* gazes (see also Lewis, 2020b). By this we mean that a topological approach changes both what is found when looking at data infrastructures; and, at the same time, how, where and to what end one asks questions and seeks to understand data infrastructures in the first place: as mechanisms, professions and practices that enact movement, while simultaneously 'holding together what has moved' (Gulson et al., 2017: 9). The latter specifically refers to the dynamic nature of the EDFacts *bordering process* (see also Newman 2006) which, in our case, occurs through the topological dynamics of data infrastructuring. In other words, as this example of border *as process* within the

data infrastructure shows, the questions raised by topological thinking, and the methodological gazes employed by the researcher, are much more *processual* in nature; for instance, (1) what are the means by which relations are formed, stretched and reformed; (2) how are proximity and distance made palpable; and (3) how are processes of topological re/de/bordering experienced?

Operationally guided by these questions, our analysis methodically builds on a qualitative content analysis of 25 policy documents related to EDFacts available online (data flow charts, descriptions of EDFact components, explanations of business rules and standards), as well as the different EDFacts online platforms and portals. From a topological point of view, such documents, platforms and portals can not only be regarded as sources for gathering information on EDFacts data, people and functions. Instead, such documents themselves *are* manifestations of the ‘labour of infrastructuring’, which means they are a central textual disposition of activities that ensure data-flowability and de/re/bordering. It is documenting as a key activity in EDFacts (see also more detailed explanation in the next section) through which the how, whom and when of data flows are explicated; for instance, how state Education Agency data is supposed to flow to the federal Department of Education. Consequently, this approach allows researchers to identity how components of the infrastructure are made relational, and how components are positioned within the infrastructure in a particular way.

To put this even more strongly, there is no component of the infrastructure without its documentation, because the document *is* creating the relation – a (topological) continuity – while simultaneously relations get ‘thicker’ the more details the documentation entails. With thickening relations, we mean that more documentation about how data activities (e.g. submission, processing) are supposed to occur, the more the activities of the various people involved in EDFacts become regulated, and the more their activities then focus on securing data flows in compliance to these regulations. Argued the other way around, the sheer number of documents on particular parts of the infrastructure (e.g. the significant number of business rule documents defining the format(s) in which data can only be submitted to EDFacts) make traceable which relations have become objects of ‘thickening labour’ or ‘infrastructural care’ (see below), and which, in turn, can be regarded as key components of the EDFacts topology.

A topological analysis of the EDFacts data infrastructure

Unlike any federal activity in US schooling until the 1950s, both the *Civil Rights Act* of 1964 and the subsequent *Elementary and Secondary Education Act* (ESEA) of 1965 clearly strengthened the federal role in securing the equalisation (and standardisation) of school access and quality (Hartong, 2021; Lewis et al., 2020). The ESEA is of particular relevance here because it implemented a systematic, standardised federal funding scheme (including Title 1, see footnote 1) for supporting students in need. Accompanying these historical developments was a multi-level evaluation and monitoring procedure, overseen by the federal Department of Education (ED). The amount of funding was calculated using a statistical formula, and computed on the basis of nationally standardised demographic and economic data (Bailey and Mosher, 1968: 49). These data themselves had to be collected in a nationally standardised way, thus helping to fabricate the first ‘topological’ space of nationally related and aligned calculation. Simultaneously, states were obliged to also transmit evaluation data for federally funded projects, ensuring ‘that effective procedures, including provision for appropriate objective measurements of educational achievement, will be adopted for evaluating at least annually the effectiveness of the programs in meeting the special educational needs of educationally deprived children’ (Bailey and Mosher, 1968: 51).²

While ESEA data initially covered only a particular funding segment of schooling, it clearly opened doors to further datafication, particularly when regulations for coupling federal funding

with reporting requirements became increasingly refined in the following decades (including the *No Child Left Behind Act* in 2001 and the *Race to the Top* program in 2009, etc.). After the 2000s, this refinement simultaneously came along with major data standardisation and interoperability initiatives, as well as technological transformations. A crucial event here was the *Statewide Longitudinal Data Systems* (SLDS) grant program, which was launched by the Institute for Education Sciences (IES) in 2005 – the statistics, research and evaluation arm of the U.S. Department of Education (<https://ies.ed.gov/aboutus>) – to support states in building standardised data-rich longitudinal school monitoring systems. Shortly after, new federal regulations were put in place in 2007, which not only required states to report certain information electronically, but also established that the federal department may take administrative action and funding withdrawal against states for their failure to submit reports (see also EDFacts, 2017: 4). It was within these broader contexts that EDFacts was launched, officially ‘to significantly reduce reporting burden and to streamline data collections currently required by the Department by bringing elementary and secondary education data in through the Annual Mandatory Collection of Elementary and Secondary Education Data for EDFacts’ (U.S. Department of Education, 2018: 7).

Since its inception in 2007, EDFacts has not only become increasingly refined and standardised (see also section ‘The rise of “new” professionals’), but the categories and quantity of data collected though EDFacts have constantly expanded. There are currently more than 100 indicators, including general information of schools, teacher and students; information on particular groups with special needs (e.g. migrants, English learners, students with disabilities, low-SES students etc.); as well as data on assessment, graduation and dropout rates, and student health and safety (see EDFacts, n.d., Meet ED data: 2).

The EDFacts infrastructure is arranged around multiple distinct yet interconnected systems. Data enters EDFacts via one of two data submission systems: (1) the *EDFacts Submission System* (ESS), which states use to submit their elementary and secondary educational data; and (2) the *EDFacts Metadata and Process System* (EMAPS), which simultaneously collects metadata, but also documents such as obligatory state data submission plans (which mark a core mechanisms of the infrastructural caring apparatus, see below).³ Upon being entered into the EDFacts submission systems, data are then duplicated into the online *Data Management System* (DMS), which is the main system designed to manage and validate submitted data. Once data have been validated and confirmed, they are then uploaded to the *Common Core of Data* (CCD), which is the ED’s primary database on public elementary and secondary education (<https://nces.ed.gov/ccd/>). From there, data are further processed into various forms for purposes of reporting, visualisation etc. These includes ‘summary statements in grant performance letters, annual reports to Congress, profiles or data display products, machine readable data files, and summary documents with key data results and visual data displays’ (EDFacts, n.d., Maximizing EDFacts data quality: 5).⁴ Processed and visualised EDFacts data do not only flow into policy information and media, but are also linked to high-stakes evaluation of grants (e.g. Title 1), while simultaneously being fed back to the states to inform their schooling policy and governance. As this general overview readily demonstrates, there are highly standardised and limited ways in which data in EDFacts can flow, including specific sets of rules that make certain practices (im)possible at each stage of the data journey.

In the following sections, we will now seek to illuminate these rules and practices in more detail, focusing first on the *whom* of infrastructuring; that is, the professionals ‘behind the scenes’ of EDFacts who are responsible for maintaining its functionality and data flow. As we will show, these professionals play an important role in the stabilisation of relations and, thus, the definition of (im)possible data movements. In other words, they regulate what is permitted inside the infrastructure and are thus constantly (re)defining its borders. Simultaneously, these professionals have been acting as key driving forces in the ongoing expansion of EDFacts, when aligning objects and

subjects outside of EDFacts to the system. This expansion not only refers to the growing amount of data produced in and about EDFacts, but also to the rising presence of EDFacts for educational actors at different levels of the schooling system, who are forced to submit themselves to the infrastructural renderings by taking active roles in ‘caring’ for the system.

The rise of ‘new’ professionals: Curating EDFacts and fostering its expansion

Over the past decade, different professionals have either become established or incorporated into EDFacts, with their main tasks being to manage, organise and support the system. The most prominent example of such roles are *data stewards* (or data stewarding offices) within the federal Department of Education, with one steward being assigned to a specific group of EDFacts data. Tasks of the data stewards include responding to state requests when submitting data, keeping up with regulations and reporting requirements, ensuring data quality pre- and post-submission and improving data usage across stakeholder groups:

The effective steward of a data file knows who is using the data, when they use it, and can anticipate what decisions will impact a nonsteward data user. [...] Stewardship is working when stewards are knowledgeable and clear about reporting requirements, but are also cooperative and collegial with nonstewarding offices that are dependent on the data. (EDFacts, n.d., Stewarding Data: 2)

Distinguishing between data stewards and non-stewarding data offices thus forms an important differentiation and also hierarchisation of actors, which, despite working in the same (federal) institution, become differently related and integrated into the enactment of the infrastructure. The stewards have effectively become a major ‘human barrier’ to possible data entry into EDFacts, by not assessing the validity of submitted data and doing edit checks but also defining an ‘acceptable level’ of errors (EDFacts, n.d., Maximizing EDFacts data quality: 4).

Alongside the data stewards, the ED has implemented an EDFacts Governing Board (EDGB), which provides a kind of ‘meta-governor’ for the whole infrastructure and ensure constantly improved data quality and flows. Therefore, it sets out to:

Set the rules, policies, and procedures related to data management and establish the lines of communication to do so . . . Provide user documentation and guidance, and the resolution of data submission issue escalations to support the submission of EDFacts data, . . . Measure and improve the timeliness, completeness, accuracy, validity, and usability of EDFacts data through a system of data quality checks. . . Expand the capacity of EDFacts users to effectively access and use EDFacts data. (EDFacts, n.d., Introduction to the US Department of Education Data Governance: 1)

At the state level, the most important EDFacts profession is the *EDFacts Coordinator*, a SEA staff member who serves as the official EDFacts contact for the federal department and who are not only responsible for the state data submission plans, but also for ensuring data submission, error correction (see next section) and federal data acceptance (U.S. Department of Education, 2019: 1). Simultaneously, EDFacts coordinators approve and monitor other data submission system users in their agency, such as the SEA *EDFacts Data Submitters* who operate ‘below’ the EDFacts Coordinator, usually as programmers or contractors. Their tasks include not only the technical submission of data files, but also the review of error reports and the submission of status reports (U.S. Department of Education, 2018: 17).

Finally, different cross-level structures and actor groups operate between the federal and state level to foster better coordination and alignment. They include (1) an online *Partner Support*

Centre, which provides training opportunities, a telephone-based support centre and a technical help desk; (2) the *EDFacts Community*, an online platform linked to the Partner Support Centre, which promotes collaboration, knowledge sharing and interaction among the EDFacts Coordinators and the larger EDFacts community (U.S. Department of Education, 2018: 18); and so-called (3) *EdTech Supporters*, who are (often) private vendors (e.g. the ESP Solution Group, AEM, escholar etc.) that offer products or support with the data submission and coordination process.

All these examples illuminate how the enactment, operation, and stabilisation of the EDFacts infrastructure relies on a network of specifically ‘designed’ professionals, who are supposed to mutually stabilise and enforce particular data practices and flows. Topologically speaking, these professionals are the means to keep the infrastructure running (i.e. receive data, use data, generate data, feed data back into the system etc.); to prevent data and data processes that do not follow the logics, norms and rules of the infrastructure (e.g. timeliness, completeness, accuracy, pass edit checks, no errors, etc.); and to help encourage the acceptance of EDFacts by its users. Such professions include certain actors from within more territorially ‘bordered’ institutional settings – for instance, the federal department or the SEAs – who become related and ordered within new hierarchies as part of their enfolding into EDFacts (e.g. between data stewarding and non-stewarding offices; or between EDFacts Coordinators and EDFacts Data Submitters). Put differently, although the topological nature of the EDFacts infrastructure enables it to ‘reach across and into’ (Allen and Cochrane, 2010) many educational institutions within territorial jurisdictions, such as the states, it nevertheless only includes and empowers *some* actors (e.g. data stewards) while excluding many others (e.g. non-stewarding offices). While this empowerment might at first glance appear to be a data-based professionalisation of certain educational actors, we will demonstrate how a closer exploration actually reveals how *all* of these professionals are being strongly ruled and constantly surveilled by the EDFacts infrastructure, its logics and norms.

Enacting, empowering and enlarging EDFacts

This section focuses on so-called *second-order activities* (Power, 2003), which are activities directed at securing and increasing the flow-ability of data in EDFacts. In particular, we refer to three significant kinds of activities around (1) *timeliness*, (2) *data error prevention* and (3) *documentation and reporting*. We distinguish these data-focused activities and abstractions from *first-order activities* (Power, 2003); these might include school community or pedagogical projects that states have undertaken with federal funding, or broader processes around student learning, but which do not ‘count’ by virtue of the EDFacts ‘rules’.

EDFacts has constituted its own space-times, and these are prominently visible in the ‘time regime’ of certain requested data activities. All professionals associated with EDFacts are ruled by strict timelines, which forces them to constantly keep track of ongoing due dates, data submission preparation and reporting requirements. Simultaneously, these due dates and data collection cycles do not follow the regular school year, but instead employ a different logic of time pressure that continuously accelerates as data are requested from the states. As noted above, states that do not submit complete and timely data or data submission plans (which is meta-data on the state’s data submission strategy) may be omitted from programmatic reports prepared for Congress, be cited for data submission failure or have federal funding punitively withheld (EDFacts, 2017: 4).

Simultaneously, to prevent ‘bad’ or error-laden data from entering the infrastructure, EDFacts has installed different ‘hard bordering’ mechanisms (EDFacts, n.d., Maximizing EDFacts data quality: 4). These include different stages of automated (technical) and human error management process to ensure submitted data are aligned to a (constantly growing) number of standards and business rules. Error reports are created on the basis of these business rules, which then

automatically prevent incorrect data from entering the next stage of the infrastructure. Errors may include format and validation errors, submission errors, or match errors, but this also applies to data which are inconsistently reported, unreported or misreported (EDFacts, n.d., Maximizing EDFacts data quality: 4). In the case of errors, SEAs are required to correct, comment on or explain data anomalies via the Data Management System (DMS), with the DMS intended to function as a real-time online communication tool between state and federal EDFacts professionals (U.S. Department of Education, 2016). Taken collectively, these processes reflect the clear presence of ‘aesthetic practices’ (Ratner and Ruppert, 2019), whereby data are *purged* of ‘inconsistencies, differences and uncertainties’ (2). While this might be interpreted as expressing at least some concern for overall accuracy, we would contest that the objective ‘truth’ of the data – that is, its alignment with the schools and schooling practices they seek to represent – is perhaps less significant to these shadow professionals than ensuring the data they handle comport with the rules and uses of EDFacts. In other words, the data fidelity that matters most is to the EDFacts business rules, rather than focusing on whether the data meaningfully depict the ‘reality’ facing teachers and students, nor (for that matter) whether these data are even useful to teaching and learning.

Finally, we cannot over-emphasise just how extensive the process of datafication is within EDFacts. Indeed, there is literally no data activity in EDFacts that is not documented or reported in some way. EDFacts hereby reflects how the attention of all participating actors is increasingly shifted towards generating data *about data*. This not only includes the state data submission plans, but also the enormous tracking lists that document the actions taken by the various state and the federal departments (and their EDFacts-related actors). These notionally inform the ongoing ‘data quality strategy cycles’ (EDFacts, n.d., Maximizing EDFacts data quality: 5), each of which is (again) followed by new regulations, standards and business rules. In sum, these second-order activities constantly (re)enact the flow-ability of data (i.e. topological continuities), while simultaneously supplementing these data with further (documentary or regulation) data; seemingly, data beget more data, *ad infinitum*. Non-aligned or ‘cracked’ spots (i.e. topological discontinuities) in the infrastructure are thus not only continuously identified, but they are also ‘fixed’ by adding further regulations, with these activities subsequently fed-back into EDFacts. At the same time, each regulation, once created, cannot be undone but only further updated and refined *within* the infrastructure. Put differently, even though the infrastructure is constantly changing, or ‘deforming’ in topological parlance (as it acquires ever-more data, new business rules, new standards etc.), the *function* of the infrastructure remains constant; that is, as a repository of data, and a means of knowing the various systems that contribute to holding EDFacts together. Topologically speaking, these constant shifts in *form* (i.e. deformations) do not change the *function* of the infrastructure, but suggest how mutable topological spaces and relations challenge the more traditional territorial spaces of school education agencies, schooling districts and classrooms (i.e. people and places).

Discussion and conclusion: EDFacts and the acts of ‘data submission’

Our purpose in this article has been to further develop topology as an analytical approach to study data infrastructures, including the primacy of data flows, and to better understand the role of emergent data-responsive ‘shadow’ professionals and practices within these infrastructures. Using the example of EDFacts, we have argued that a topological approach is not only useful, but also necessary, to understand how processes of alignment, normalisation and governance take place when facets of schooling become datafied. For example, EDFacts has incentivised significant national standardisation and compliance to federally driven standards, norms and regulations across states (see also Hartong, 2021; Lewis et al., 2020; Savage and O’Connor, 2019), yet without the federal

government directly ‘ruling’ the system, nor the states officially ceding power or responsibility to federal agencies.

A key purpose of this article has been to illuminate how a topological approach can be used not only as a theoretical set of metaphors, but rather as an *empirical* device, which necessitates a shift in both *methodological* and *analytical* gazes. By this we mean that a topological approach changes both what is found when looking at data infrastructures; and, at the same time, how, where and to what end one asks questions and seeks to understand data infrastructures in the first place: as mechanisms, professions and practices that enact movement. Rather than focusing on changing power distributions between or within territorial spaces (e.g. the national, the state/subnational, the individual government departments), topological thinking instead invites us to explore the governing spaces that emerge and operate through different modes of bordering and flows of data, as in the case of EDFacts. As our findings show, the governmental power of alignment and normalisation has increasingly shifted to the infrastructure itself, particularly in its logic of producing flawless data; here, defined as adherence to business rules and standards, rather than achieving objective accuracy. Such logics are reminiscent of Porter’s (2012) concern for ‘funny numbers’, where having ‘clean’ data is perhaps more important than having accurate data, in the sense that they might reflect the reality they putatively represent, or even provide notionally useful insights.

These data infrastructures, however, don’t just provide new spaces of governance but also interact with existing territorial spaces. Indeed, such territorial spaces are still relevant, insofar as they make the topological relations of the infrastructure possible, and these spaces are authorised to use the data and analyses derived from EDFacts (see also Lewis, 2020b). In short, these are still the spaces where many political decisions are made. Nevertheless, the role of the various (shadow) professionals, and their concern with notionally ‘second-order’ activities, also reflect how establishing and maintaining topological spaces like EDFacts is exceptionally labour intensive. The more that data move, in terms of their quantity and frequency of access, the more that bordering investments are required (e.g. developing new business rules), and the more professional roles like data stewards are both responsible and necessary for securing these borders. This also means, however, that ‘professionalisation’ in EDFacts means having to continuously increase the power of non-human business rules, standards, timeliness and documentation loops. Despite the internal hierarchisation of positions (e.g. the key gatekeeper role of the data stewards to ‘let data in’), this pressure is equally exerted upon all EDFacts users; that is, to adhere to the specific business rules and regulations of EDFacts, but also, more generally, to profess a data-responsive disposition and approach to understanding and practising schooling (Holloway, 2019; Lewis and Holloway, 2019).

Put differently, it is following protocols, timelines and regulations, rather than exhibiting concern for the actual content of the data and what it implies for teaching and learning (e.g. how well are schools doing; does student learning improve?), that becomes the *key profession* of the data stewards, error managers and data commissioners within EDFacts. In this sense, the etymological basis of the verb ‘to steward’ – that is, to look after or manage another’s property – is particularly apt here: to steward data is to serve data, and thus ensure the stipulations of EDFacts are maintained at all times. Additionally, even if the rules and documents may initially be authored by some of these human actors, these technical standards are then absorbed and suffused throughout the system, ultimately removing identifiable ‘human’ traces. This also sheds new light on the critique that datafication means professionalising data experts while de-professionalising others, such as teachers. Based on our analyses of EDFacts, we would instead argue that educators and data stewards are *equally* de-professionalised by the need to uncritically respond to and service data. In effect, the key characteristic of teaching professionals and shadow professionals alike are their *acts of data submission* to EDFacts: by way of providing data to EDFacts and, at the same time, by positioning themselves as wholly responsive to the technical aspects of the infrastructure and its datafied renderings of schooling.

Finally, our findings clearly showed how data infrastructures such as EDFacts continuously enable powerful self-reinforcing dynamics of datafication. Regardless of the specific problem or challenge, the response is invariably adding more and ‘better’ data, supported by additional shadow professionals and business rules to help ensure its quality. This not only extends the infrastructure but also, paradoxically, compresses it through intensifying monitoring and accountability processes. We wonder what might happen if one takes such data-venerating and fetishising logics to their extreme conclusion: if data are so important as to warrant data stewards, then what is needed to ensure the quality of the data stewards themselves? Data-steward *stewards*, or error-manager *managers*, perhaps? And will this oversight be trusted to people, or will the designers of EDFacts seek to overcome potential human subjectivity and poor judgement by reverting to algorithmic oversight? One cannot help but be reminded of the well-worn question posed by the poet Juvenal in his *Satires*: *Quis custodiet ipsos custodes?* That is, *Who watches the watchers?* Moreover, we would question whether systems such as EDFacts even make it possible for teachers and other users to dispute the veracity and centrality of data, or the new governing regimes constituted through topological relations and spaces. Rather than focusing on hypothetical overseers of data stewards, we will close here by arguing that it is precisely *against* such a blind insistence on trusting, producing and servicing data – often at the expense of meaningful human experience and professional judgement – where we all need to be most watchful.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Australian Research Council [grant number DE190101141] for Steven Lewis and by the German Research Foundation [grant number HA 7367/2-1] for Sigrid Hartong.

ORCID iD

Steven Lewis  <https://orcid.org/0000-0002-8796-3939>

Notes

1. In the United States, ‘Title I is the largest federally funded educational program. The program provides supplemental funds to school districts to assist schools with the highest student concentrations of poverty to meet school educational goals. . . . Schools must make adequate yearly progress on state testing and focus on best teaching practices in order to continue receiving funds’ (<https://definitions.uslegal.com/t/title-1-school/>).
2. A few years later, policy actors who had initially contributed to the empowerment of the federal department also started to promote the implementation of a nationally standardised, yet sample-based, performance assessment. This became the *National Assessment of Educational Progress* (NAEP).
3. While ESS and EMAPS operate in tandem, they are serviced by separate EDFacts coordinators and working groups.
4. Additionally, the CCD collection flows into the sampling framework for the National Assessment of Educational Progress (NAEP) (EDFacts, 2017: 2).

References

- Allen J and Cochrane A (2010) Assemblages of state power: Topological shifts in the organisation of government and politics. *Antipode* 42(5): 1071–1089.
- Amoore L and Piotukh V (2015) Life beyond big data: Governing with little analytics. *Economy and Society* 44(3): 341–366.

- Bailey SK and Mosher EK (1968) *ESEA: The Office of Education administers a law*. Syracuse, NY: Syracuse University Press.
- Billé F (2018) Skinworlds: Borders, haptics, topologies. *Environment and Planning D: Society and Space* 36(1): 60–77.
- EDFacts (2017) EDFacts discussion points. Available at: <https://www2.ed.gov/about/inits/ed/edfacts/edfacts-discussion-points.docx> (accessed 6 June 2020).
- EDFacts (n.d.) Introduction to the US Department of Education Data Governance. Available at: <https://www2.ed.gov/about/inits/ed/edfacts/edfacts-edgb-data-governance.pdf> (accessed 6 June 2020).
- EDFacts (n.d.) Maximizing EDFacts data quality: A comprehensive approach (2020) Available at: <https://www2.ed.gov/about/inits/ed/edfacts/edfacts-data-quality-process-overview.pdf> (accessed 6 June 2020).
- EDFacts (n.d.) Meet ED data. Available at: <https://www2.ed.gov/about/inits/ed/edfacts/meet-ed-data.pdf> (accessed 6 June 2020).
- EDFacts (n.d.) Stewarding data: Subject matter expertise and data governance. Available at: <https://www2.ed.gov/about/inits/ed/edfacts/edfacts-data-stewarding-overview.pdf> (accessed 6 June 2020).
- Grek S (2009) Governing by numbers: The PISA ‘effect’ in Europe. *Journal of Education Policy* 24(1): 23–37.
- Grommé F and Ruppert E (2019) Population geometries of Europe: The topologies of data cubes and grids. *Science, Technology, and Human Values* 45(2): 235–261.
- Gulson K, Lewis S, Lingard B, et al. (2017) Policy mobilities and methodology: A proposition for inventive methods in education policy studies. *Critical Studies in Education* 58(2): 224–241
- Gulson K and Sellar S (2019) Emerging data infrastructures and the new topologies of education policy. *Environment and Planning D: Society and Space* 37(2): 350–366.
- Hartong S (2016) Between assessments, digital technologies and big data: The growing influence of ‘hidden’ data mediators in education. *European Educational Research Journal* 15(5): 523–536.
- Hartong S (2018) Towards a topological re-assemblage of education policy? Observing the implementation of performance data infrastructures and ‘centers of calculation’ in Germany. *Globalisation, Societies and Education* 16(1): 134–150.
- Hartong S (2021) The power of relation-making: Insights into the production and operation of digital school performance platforms in the US. *Critical Studies in Education* 62(1): 34–49.
- Hartong S and Piattoeva N (2019) Contextualising the datafication of schooling: A comparative discussion of Germany and Russia. *Critical Studies in Education*. Epub ahead of print 20 May 2019. DOI:10.1080/17508487.2019.1618887.
- Holloway J (2019) Teacher evaluation as an onto-epistemic framework. *British Journal of Sociology of Education* 40(2): 174–189.
- Kitchin R (2014) Big Data, new epistemologies and paradigm shifts. *Big Data and Society* 1(1): 1–12.
- Landri P (2018) *Digital governance of education: Technology, standards and Europeanisation of education*, London, Bloomsbury Academic.
- Lawn M (2013) Introduction: The rise of data in education. In: LAWN, M (ed.) *The rise of data in education systems: Collection, visualisation and use*. Oxford: Symposium, 7–10.
- Lewis S (2020a) *PISA, policy and the OECD: Respatialising Global Educational Governance Through PISA for Schools*. Singapore: Springer Nature.
- Lewis S (2020b) The turn towards policy mobilities and the theoretical-methodological implications for policy sociology. *Critical Studies in Education*. Epub ahead of print 14 August 2020. DOI:10.1080/17508487.2020.1808499.
- Lewis S and Holloway J (2019) Datafying the teaching ‘profession’: Remaking the professional teacher in the image of data. *Cambridge Journal of Education* 49(1): 35–51.
- Lewis S, Sellar S and Lingard B (2016) ‘PISA for Schools’: Topological rationality and new spaces of the OECD’s global educational governance. *Comparative Education Review* 60(1): 27–57.
- Lewis S, Savage GC and Holloway J (2020) Standards without standardisation? Assembling standards-based reforms in Australian and US schooling. *Journal of Education Policy* 35(6): 737–764.
- Lingard B (2011) Policy as numbers: Ac/counting for educational research. *The Australian Educational Researcher* 38(4): 355–382.

- Lury C, Parisi L and Terranova T (2012) Introduction: The becoming topological of culture. *Theory, Culture and Society* 29(4-5): 3–35.
- Lury C, Parisi L and Terranova T (2013) Interview with Celia Lury, Luciana Parisi and Tiziana Terranova on topologies. Available at: <http://theoryculturesociety.org/interview-with-celia-lury-luciana-parisi-and-tiziana-terranova-on-topologies/> (accessed 6 June 2020).
- Martin L and Secor AJ (2014) Towards a post-mathematical topology. *Progress in Human Geography* 38(3): 420–438.
- Mezzadra S and Neilson B (2012) Between inclusion and exclusion: On the topology of global space and borders. *Theory, Culture and Society* 29(4-5): 58–75.
- Newman D (2006) The lines that continue to separate us: Borders in our ‘borderless’ world. *Progress in Human Geography* 30(2): 143–161.
- Ozga J (2016) Trust in numbers? Digital education governance and the inspection process. *European Educational Research Journal* 15(1): 69–81.
- Piattoeva N and Saari A (2019) The infrastructures of objectivity in standardised testing. In: Maddox B (ed.) *International Large-Scale Assessments in Education: Insider Research Perspectives*. London: Bloomsbury Academic, 53–68.
- Porter T (2012) Funny numbers. *Culture Unbound* 4: 585–598.
- Power M (2003) Evaluating the audit explosion. *Law and Policy* 25(3): 185–202.
- Prince R (2016) The spaces in between: Mobile policy and the topographies and topologies of the technocracy. *Environment and Planning D: Society and Space* 34(3): 420–437.
- Ratner H and Ruppert E (2019) Producing and projecting data: Aesthetic practices of government data portals. *Big Data and Society* 6(2): 1–16.
- Robertson SL (2011) The new spatial politics of (re)bordering and (re)ordering the state-education-citizen relation. *International Review of Education* 57(3-4): 277–297.
- Ruppert E (2012) The governmental topologies of database devices. *Theory, Culture and Society* 29(4-5): 116–136.
- Savage GC and O’Connor K (2019) What’s the problem with ‘policy alignment’? The complexities of national reform in Australia’s federal system. *Journal of Education Policy* 34(6): 812–835.
- Scheel S (2020) Biopolitical bordering: Enacting populations as intelligible objects of government. *European Journal of Social Theory*: 1–20. Epub ahead of print 26 January 2020. <https://doi.org/10.1177/1368431019900096>.
- Takayama K and Lingard B (2019) Datafication of schooling in Japan: An epistemic critique through the ‘problem of Japanese education’. *Journal of Education Policy* 34(4): 449–469.
- U.S. Department of Education (2016) EDFacts Technology and Support Services (ETSS) II. DMS State User Guide. Available at: <https://www2.ed.gov/about/inits/ed/edfacts/dms-state-user-guide.pdf> (accessed 6 June 2020).
- U.S. Department of Education (2018) *EDFacts Workbook SY 2018-19*. Available at: <https://www2.ed.gov/about/inits/ed/edfacts/eden/18-19-workbook-15-0.pdf> (accessed 6 June 2020).
- U.S. Department of Education (2019) EDFacts Submission System (ESS) User Guide. Available at: <https://www2.ed.gov/about/inits/ed/edfacts/eden/ess/ess-user-guide-v16-0.pdf> (accessed 6 June 2020).
- U.S. Department of Education (2020) EDFacts. Available at: <https://www.ed.gov/open/plan/edfacts> (accessed 14 February 2020).
- Van De Oudeweetering K. and Decuyper M (2019) Understanding openness through (in)visible platform boundaries: A topological study on MOOCs as multiplexes of spaces and times. *International Journal of Educational Technology in Higher Education* 16(1): 1–30.
- Williamson B (2017) *Big data in education: The digital future of learning, policy and practice*. Los Angeles: Sage.

Author biographies

Steven Lewis is a Senior Research Fellow and Australian Research Council (ARC) DECRA Fellow at the Institute for Learning Sciences and Teacher Education at Australian Catholic University. His current research project, entitled 'Globalising school reform through online teacher professional learning' (2019-2022), investigates how schooling reform, teacher practice and educational governance is being reshaped by new forms of international evidence, data infrastructures and relational policy spaces, including online learning communities such as the OECD's PISA4U. His most recent monograph is *PISA, Policy and the OECD: Respatialising Global Educational Governance Through PISA for Schools* (2020, Springer).

Sigrid Hartong is a postdoctoral research fellow at the Department of Education at the Helmut-Schmidt-University Hamburg, Germany. She holds a PhD in Sociology (University of Bamberg) and a Habilitation degree in Education Science (Helmut-Schmidt-University Hamburg). Her main scholarly interests include global-local educational reform dynamics, both from single-case and international comparative perspectives, as well as the growing datafication and digitalisation of education policy and practice, particularly visible in the rise of (big) data infrastructures and mobilities. Her previous research explored policies on educational standardisation in Germany and the US, including the implementation of curriculum standards, standardised assessments and new monitoring procedures.