Learning, Media and Technology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cjem20

Interoperability and data standards in the K-12 education sector: intersections with data justice

Molly S. Stewart, Elizabeth Pier, Dan Ralyea & Andrew Rice

To cite this article: Molly S. Stewart, Elizabeth Pier, Dan Ralyea & Andrew Rice (2023): Interoperability and data standards in the K-12 education sector: intersections with data justice, Learning, Media and Technology, DOI: 10.1080/17439884.2023.2190121

To link to this article: https://doi.org/10.1080/17439884.2023.2190121

Published online: 15 Mar 2023.

LEARNING, MEDIA AND TECHNOLOGY
https://doi.org/10.1080/17439884.2023.2190121

Interoperability and data standards in the K-12 education sector: intersections with data justice

Molly S. Stewart [a], Elizabeth Pier [b], Dan Ralyea[c] and Andrew Rice[b]

[a]School of Education, Indiana University, Bloomington, IN, USA; [b]Education Analytics, Madison, WI, USA; [c]South Carolina Department of Education, Columbia, SC, USA

ABSTRACT

This paper examines the concept of open-source data interoperability in the United States' K-12 education domain, specifically addressing the implications of interoperability for data justice. The term 'data justice' is a relatively recent coinage; the framing and meaning of this term are still evolving, and it has not yet been applied in the educational domain. Building on the nascent research and theorizing in the fields of both data justice and interoperability in educational contexts, this paper provides an overview of the current state of this intersection. Additionally, the authors draw on their direct experiences implementing interoperability initiatives in several U.S. states in order to build a foundational understanding of the risks and opportunities for data justice in the realm of interoperability and data standards in education. The paper concludes with a call for more research to be completed on this complex sociotechnical topic.

ARTICLE HISTORY
Received 31 May 2022
Accepted 2 March 2023

KEYWORDS
Interoperability; data justice; primary education; secondary education; data standard

Introduction

> Interoperability is the seamless, secure, and controlled exchange of data between applications. Interoperability allows data to easily flow among applications that are developed for different purposes using a standardized vocabulary, structure, and cadence. (State Educational Technology Directors Association 2018)

This paper examines the concept of open-source data interoperability in the United States' K-12 education domain, specifically addressing the implications of interoperability for data justice (and/or injustice). The term 'data justice' (Dencik et al. 2019) is a relatively recent coinage; the framing and meaning of this term are still evolving, and it has not yet been applied in the educational domain (although many scholars of education have addressed data justice issues without using the term itself; e.g., Clutterbuck 2022; Clutterbuck, Hardy, and Creagh 2021; Eynon 2013). In this paper, we incorporate prior framing of the concept while also finding it useful to further expand the meaning of this term as we examine specific issues, potential consequences, and future opportunities of interoperability in U.S. K-12 education.

Interoperability of data systems has become the norm in home and business technologies (e.g., Noura, Atiquzzaman, and Gaedke 2019), healthcare (e.g., Gordon and Catalini 2018), and other industries that rely on real-time data updates for essential operations and decision making. However, despite much discussion in the education sector regarding similar needs for data (e.g., Piety 2019), it has had a much slower uptake in that sector due to a combination of federal and state policies (e.g., Anagnostopoulos, Rutledge, and Bali 2013), local and state organizational cultures that have been shaped by decades of this policy interplay (e.g., Weatherly and Lipsky 1977), and constraints

CONTACT Molly S. Stewart stewarmo@indiana.edu School of Education, Indiana University, Rm 4208, 201 N. Rose Ave., Bloomington, IN 47405, USA

surrounding proprietary systems owned by private vendors. Federal and state privacy laws and accountability requirements are among the most powerful of these policy constraints, and Corbeil, Corbeil, and Khan (2019) provide a useful and concise discussion of common challenges within education organizations, including siloed data sources and non-standard data formats.

Due in large part to the slow uptake and above challenges, interoperability has received little attention in the K-12 education research literature, with Gulson and Sellar's (2019) discussion of the National Schools Interoperability Program in Australia being the exception (as well as additional related articles; Sellar 2017; Sellar and Gulson 2019). A few scholars have also noted either the general lack of its use in K-12 education (e.g., Mandinach and Schildkamp 2021; Pangrazio, Selwyn, and Cumbo 2022) or the use of proprietary interoperability among private companies (e.g., Perrotta et al. 2021), but more work is needed that focuses primarily on interoperability technologies and how the design and implementation of these technologies can influence K-12 education. This article aims to fill this gap, and specifically addresses the ways in which *open-source* interoperability has implications for data justice.

Interoperability refers to the integration of data systems using secure connections to share data either unidirectionally or bidirectionally. These secure connections are made through APIs (application program interfaces), which are externally accessible interfaces on top of the source system. The output of an API may use an internal data schema or may be aligned to an open data standard; standardization of output aligned with an open data standard is the key to open-source interoperability. With system vendors producing API output aligned with an open data standard, education agencies can choose which systems to integrate and which data to exchange (and in which direction). The open data standard, which requires that data flow between systems are in expected and standardized formats (e.g., a telephone number is 10 digits and does not include dashes or spaces), allows data users (i.e., education agencies) to understand exactly what format exchanged data will take.

In this paper we draw on our experiences implementing and supporting local, state, and multilevel interoperability initiatives using the Ed-Fi Data Standard (Ed-Fi Alliance 2023) in several U.S. states, including California, Colorado, Indiana, New Mexico, and South Carolina. However, due to our on-the-ground involvement in these implementations as project and technical staff – not as researchers – we generally do not refer to specific examples from specific locations in order to maintain the privacy of actors involved in those implementations. Our roles as practitioners in various educational technology settings (public and non-profit) uniquely position us to discuss and examine relevant issues 'in the spirit of positively modifying the ecology of practices in which we are situated' (Sellar 2015, 774; quoted in Decuypere 2021).

Literature review

In school and local education agency sites, there is a persistent observation of 'siloed' data sources that make it difficult for educators and leaders to get a multi-faceted and current picture of how individual students or student groups are progressing on various measures of academic and behavioral progress. In contrast, studies of data mining and machine learning, especially by for-profit vendors with access to big data gleaned from students' and educators' interactions with their products, focus on the problem of too much data, lack of consent or knowledge about its collection, and the use of such data for undemocratic purposes. In this paper, we focus on technologies (interoperability, data standards, and APIs) that are primarily used by for-profit education technology companies but which have the potential to address issues of student and family agency, equitable educator decision-making and resource allocation, and accountability of educational systems to socially just disciplinary expectations (if yet unrealized), all of which we consider to be under the umbrella of 'data justice.'

The following sections review the data justice literature as well as two sets of tensions that have arisen related to the concept of data justice specifically in K-12 education technology (although the education research literature does not use the term 'data justice' outright). The first set is the tension between surveillance (or 'dataveillance'; Clarke 1988) in schools and educator arguments in favor of using data to understand 'the whole child' or a '360 degree view' of a student (e.g., Krachman, LaRocca, and Gabrieli 2018). For example, Jarke and Breiter's (2019) editorial on 'The datafication of learning' discusses increased data use in schools as a juxtaposition of multiple consequences, including increased transparency as well as concerns about privacy and control. These issues are also related to the growing role of private companies in creating data-related products versus the role of local educators in making data-related decisions. The second set is the tension between increasingly advanced technical solutions regarding the organization and movement of data and the continued use of simple, manual data collection and analysis by educators. The review of these tensions shows how the complex web of technological developments and for-profit companies has changed (and will likely continue to change) the landscape of K-12 education (e.g., Gulson and Sellar 2019). Following this review, we will then discuss ways in which the particular technologies of interoperability and open-source data standards may be able to mitigate some of the potential or actual harms involved in the tensions discussed.

Data justice: scope and prior work

The term *data justice* was coined and introduced to academic theory-building, respectively, in the mid-2010s by the now-defunct U.S. organization Data Justice and by Dencik, Hintz, and Cable (2016). The concept surfaced specifically as a way to mobilize thinking and action around the ways in which big data, particularly those data that are 'digitally enabled … [by] the infrastructures of our everyday technologies' (Dencik, Hintz, and Cable 2016, 9; see also Leonelli 2019), are used for surveillance and control in society. However, from our position as education researchers, technologists, and leaders, we see the concept of data justice as relevant to more types of data than only digitally enabled 'big data.' In the area of education technologies used by and for K-12 education agencies, legal limits on the collection and sharing of digitally enabled data, particularly those data associated with a particular individual, as well as the widespread use of student data for both operational decision-making and state-driven accountability programs, result in a specific context that could benefit from the application of the concepts of data justice. Additionally, in partial contrast to Kitchin's (2014) discussion of the 'data revolution,' which includes 'a new generation of data analytics designed to cope with data abundance as opposed to data scarcity' (xv), K-12 education systems in the United States, and likely elsewhere, are somewhere in the middle of abundance and scarcity in terms of (quality) data points, management, and analytical capabilities. That is, the data justice issues arising from 'big data' are largely different than those specific education records (e.g., attendance, disciplinary events, grades, test scores) commonly collected and used in the K-12 education sphere (e.g., Lindh and Nolin 2016).

An additional consideration related to previous literature is that while education research in the context of the U.S. K-12 system has certainly examined the practices, benefits, and consequences of data utilization, prior studies have largely been divorced from discussions of technologies and how these impact the quality, availability, and use of data (Mandinach and Schildkamp 2021). Even when research does include a focus on technology, studies of educational contexts have observed that technology used for data use have been 'relatively unsophisticated' (Selwyn 2016, 63; see also Cho and Wayman 2015). However, the growing use of interoperability technologies is, in some places, beginning to change the 'simplistic' approach to data infrastructures for some education agencies and their stakeholders. We see the application of core data justice ideas to the role of interoperability technology in education, particularly related to the creation, storage, flow, and accessibility of data as a necessary next step to understanding the potential promise and dangers of interoperability for various data justice implications.

360 Degree view of a child: surveillance, holism, or both?

As noted above, surveillance or 'dataveillance' (Clarke 1988) is one of the foundational and consistent themes in the data justice and related literatures (e.g., Dencik, Hintz, and Cable 2016; Landwehr, Borning, and Wulf 2021; Taylor 2017). However, what is meant by 'surveillance' in practice can vary. Williamson (2016) examines how the use of predictive analytics algorithms by large companies can 'pre-empt' the educational futures of students in problematic ways. As discussed in Lindh and Nolin (2016), the privacy agreement for Google Apps for Education has raised concerns about both 'student data' and 'student information' – the difference between these is not quite clear, but seems to be generally educational/personal data points versus online behavior information collected by the applications (e.g., student page visits/click data). Lindh and Nolin (2016), though focusing primarily on Google's use of online behavior information for marketing purposes, also touch on the use of student-app interaction data in instruction, such as the ability to see student interactions with writing assignments as they are being created and edited.

This observation points to the flip side of this tension: educators find behavior information about students to be useful in understanding the learning process and in figuring out where students are not engaged. The increasingly popular educational approaches focusing on 'early warning' and 'the whole child' or a '360 degree view' of a student relies on multiple points of data (in the case of the 'whole child' approach, these data are regarding not just academics but behavioral engagement and even health). Early warning systems that use multiple points of data to identify students who may need extra supports or interventions have remained popular since their introduction in the mid-2000s (e.g., Allensworth and Easton 2007; Balfanz, Herzog, and Mac Iver 2007). Support for the whole child approach has come from academia (e.g., Murray, Hurley, and Ahmed 2015; Wikman, Allodi, and Ferrer-Wreder 2022), professional organizations (e.g., Krachman, LaRocca, and Gabrieli 2018), private foundations (e.g., Chan Zuckerberg Initiative 2023), and system vendors (e.g., Education360 2023; Ehlers 2020; Karapetyan 2018), and multiple professional conferences on the topic now exist (e.g., Educating the Whole Child Summit 2023; the Whole Child Conference 2023; Whole Child Summit 2023).

Several scholars discuss this tension (e.g., Charteris 2022; Jarke and Breiter 2019; Phelps and Santo 2022; Rafalow and Puckett 2022). Hawn Nelson et al. (2020) also provide recommendations for data use best practices that consider potential negative consequences such as racial or socioeconomic profiling – two demographic groups who already experience over-surveillance in many social settings (e.g., Archibald 2022; Eubanks 2018; Iwama et al. 2022). Taylor (2017) notes that there is a disconnect in much of the literature on data-related technology: 'The frameworks we currently use either emphasize risk and harm, or argue for making data and the power to analyze them as broadly accessible as possible. The task of reconciling these perspectives is politically and theoretically huge' (12). Selwyn, Pangrazio, and Cumbo (2022) take on this task and provide a reframing of how data ('Big' and otherwise) can play a useful role – not the only role – in pedagogy.

While research has established some best practices for educators, a continuing concern is that the private companies involved in facilitating the '360 degree view' both may not be interested in best practices for educational equity and also are in many ways beyond the control of the local educational leaders using their products (e.g., Komljenovic 2021). Eynon (2013) provides an insightful editorial on the justice issues and potential consequences for equity related to the use of data (including Big Data) in educational settings, and Williamson (2015), Williamson et al. (2022), and Kerssens and van Dijck (2022) discuss the competing issues of governance between education agencies and for-profit companies that are providing data-related products and services. Thus, the tensions noted here, both regarding surveillance/educational supports and educator/for-profit governance, are far from resolved.

Evolving technologies and human/organizational capacity

This section reviews literature that has examined aspects related to interoperability and organizational capacity. While technologies around movement and use of data have become extremely sophisticated in some ways (particularly in the private sector), education organizations have advanced their data infrastructures more slowly, particularly around interoperability. In their

2021 review of common misconceptions related to data-based decision making, Mandinach and Schildkamp highlighted 'lack of interoperability among the silos of data and teachers' knowledge of how to triangulate across data sources' (n.p.) as two of the key issues in the overarching misconception related to the role of technology in data-based decision making. Piety's (2019) review of actionable data use highlights infrastructure as the central component allowing educators to achieve actionable data usage in practice.

Pangrazio, Selwyn, and Cumbo's (2022) study of data infrastructures in Australian secondary schools provides evidence that the actual implementation of interoperability is complex, involving multiple proprietary systems and varying levels of interest in achieving interoperability. Their study also reveals how, when vendors of proprietary products do not buy into the interoperability programs of education agencies, significant human labor is required for makeshift solutions (e.g., manual data exports and imports from one system to another system).

We also see a limited understanding of data interoperability being advanced in some literature. For example, Hartong (2016) discusses the requirement of interoperability in 'Big Data' collection but does not also recognize that it can be useful for purposes other than Big Data – including the purpose of breaking down the data silos within individual education agencies. Selwyn (2022) bridges the observations of Hartong (2016) and Pangrazio, Selwyn, and Cumbo (2022), drawing on additional infrastructure studies to address the labor-intensive work of data infrastructures, as opposed to automatic and massive collections of Big Data. An additional point not noted in any of the literature is the specific set of work and processes needed to move from manual data movements to directly connected flows of data. The process to set up direct flows of data is not automatic – it takes work – but the result is that data can then flow without (much) human intervention or labor. Common conceptions of Big Data as flowing automatically and without user consent often fail to mention that this occurs due to ownership of those technologies by companies, and therefore the companies can set up the automatic data flows (Hartong 2016). In the K-12 educational landscape, most data systems have different owners (public or private), and connections among those systems take both work and consent by the *data* owners (the education agencies) as opposed to the *system* owners (the vendors).

Selwyn's (2022) study of data use in secondary schools provides additional evidence of how on-the-ground data work is often done outside of any existing technical 'infrastructures' at all, but rather with manual collection and manipulation of 'small data' points for very specific purposes. Selwyn asks, 'why there might be 'good' organizational reasons for 'bad' school data' (109). Considering the difficulties noted above in aligning and connecting the myriad of data systems in an education agency, paired with significant evidence from the data-based decision making literature on the difficulty of changing educator data use practices (i.e., convincing staff to utilize a new system; Cho and Wayman 2014), the findings from Selwyn (2022) are unsurprising. Staff will use the tools that they have and are familiar with in order to analyze data in the ways that are needed. Unfortunately, these ways are often time- and effort-consuming, which takes away from educators' primary role – working with students directly (Selwyn 2021).

Finally, Clutterbuck's (2022) study of Queensland's state-wide student information system discusses various issues related to data infrastructure; however, the study does not always make clear whether an issue is based in the technology itself or in the organizational and human user decisions around the technology (i.e., data governance). Technology and data governance are key foundations for organizations to be able to utilize technology in ways that are appropriate according to policies and ethics. Clutterbuck's criticism of the 'source of truth' narrative is an example of this: 'single source of truth' is a well-established concept in software engineering that reduces data replication and human error (e.g., Pang and Szafron 2014). For example, if a school uses a learning management system (LMS) to record grades, but report cards are created from a different system (usually the student information system), the organization must decide which of these two are the source of truth for report card grades. Ensuring that an organization's data system has checks and balances for data errors and inconsistencies is established best practice in the field, and Clutterbuck's study does not recognize that 'source of truth' is more than a political narrative.

Discussion: data justice implications for interoperability with open data standards

Metadata and context

Data are recorded summarizations by an observer of some (possibly unknowable) state of the world. Because any particular state of the world is infinitely granular, the summarization component of this definition is key. It implies that any piece of data is inherently not a direct or 'true' representation of the state of the world. This has been termed the *representation view* of information (McKinney and Yoos 2010). On the other hand, this definition also implies that any summary of an observation (the piece of data) is 'objectively true' in the sense that that particular observer interpreted the observable event in a bounded context. Thus, the recording of the state of the world then becomes part of the state of the world (in a record/dataset) and so therefore runs into the same complexities of the data collection itself.

This gives rise to data collections about data and its collection (or *metadata*; Kitchin 2014). Metadata are necessary to understand the observed piece of data and its interpretive limitations. However, especially for observations that are qualitative in nature, metadata is inherently incomplete, as the complete state of the world for that observation is too complex to collect and record. Additionally, research has shown that observer recall of what they have observed is often incomplete or inaccurate (e.g., Lavis and Brewer 2017), which of course has more problematic implications for observation data being used for decision making. However, despite these limitations, the collection of a consistent subset of metadata is not only useful but necessary for understanding the observations themselves, and may even be able to provide some checks on observer unreliability, in the sense that the collection of metadata enables analysis of patterns within it. Interpretation of any data requires relevant context for the purpose at hand, otherwise the interpretation that will be used will reflect the beliefs and understandings (flawed or not) of the person interpreting the data, as opposed to the perceptions and observations of the original observer/recorder of the data. If we believe that the reason to gather and utilize data in education is to understand aspects of the realities of education that go beyond what is directly observable by a classroom teacher or other staff member, then we must recognize the importance of including contextual metadata in our design and use of data-based activities.

Scholars in the field of library and information sciences have begun to address similar issues regarding the interplay of historical/contextual representation and social justice. For example, Jaffe (2020) describes an apparent dichotomy in how the field has begun prioritizing 'big data scenarios' (435) over human implementers in the use of metadata, meaning that the categories of metadata that are prioritized are those needed for system integration, not those that are most relevant human users of metadata. This observed dichotomy appears to separate the requirements of interoperability from human users. However, in our experience with K-12 education agency implementation and utilization of interoperability, these two aspects are (or should be) intimately connected. On the other hand, the metadata currently collected in the K-12 education domain is utilized by human users far less than it could be; this low usage is likely due to a combination of siloed operations (technology and instruction) as well instructional leaders' awareness of and ability to access metadata. The integration of data systems through interoperability allows for granular data and metadata collected by the original system to remain linked to commonly used data points.

Consequentially, utilizing metadata also introduces additional privacy and security concerns for educational organizations. The benefit of individualizing or contextualizing educational data comes with the risk of increasing the amount and detail of personally identifiable information in datasets. When these datasets are transmitted beyond the local education agency (e.g., for state accountability reporting), the risk of misuse of such data heightens. Education practitioners and policymakers must weigh the benefits of having more accurate, granular, and contextualized datasets with the risks of such datasets being used for purposes that are antithetical to student equity and achievement. Ralyea and Rice (2022) have explored student-focused use cases related to metadata and context; more work is needed in order to fully understand the possibilities and limitations in this area.

Power dynamics among vendors, states, and local agencies

The standardized foundation of open-source interoperability tools (APIs and data standards) has several implications for education agencies. First, system vendors have less power to control which products are used by agencies. Connections using APIs with standardized output are interchangeable, meaning that if two system vendors providing similar features both have standardized API integrations, an education agency can choose which vendor best fits their needs and can more easily change vendors without concern that their data is locked in a proprietary data warehouse owned by the original vendor. The increased use of an open data standard provides additional market leverage to education agencies; agencies can include specific interoperability requirements in purchasing bid requests and contracts (Dexter, Francisco, and Luna 2021). Using an API with standardized data formats also allows education agencies to choose among data storage, data analysis, and visualization vendors that also integrate with the API. As Landwehr, Borning, and Wulf (2021) point out, open-source interoperability can level the playing field among education agencies and vendors (including small vendors who may be competing in a heavily Big-Ed-Tech dominated marketplace): 'APIs to support adversarial interoperability and portability are key to enabling a flourishing ecosystem of different applications that can function together and that allow end users to move to different providers. … In general, interoperability counteracts overdependence on the part of users on service providers and reduces the possibilities of cutting off innovative competitors' (10). The danger of education technology markets being dominated by 'Big EdTech' companies has been discussed by Williamson (2022).

Finally, interoperability can be initiated and facilitated by different stakeholders in the education domain. While policies mandating interoperability often come from a state or national education agency for the purposes of accountability reporting (e.g., Gulson and Sellar 2019), local agencies can also work directly with their contracted vendors to implement interoperability for local purposes (i.e., data-based decision-making). The use of open data standards here also has the potential to decrease the power imbalance between state and local education agencies (compare Anagnostopoulos and Bautista-Guerra 2013). Although interoperability has been slowly gaining traction in the K-12 education domain since the early 2000s (Gulson and Sellar 2019), challenges continue to constrain current interoperability implementations. These challenges are diverse and often in flux, given changing technical and policy developments at multiple governmental levels. Thus, the specific details of 'interoperable' systems can vary and have varying implications for educators, for students, and for data justice concerns (Kitchin 2014).

Access and visibility of data for student subgroups

The first technical requirement for interoperability is the use of a common data standard, which enables different systems to 'speak the same language' as data is transmitted or exchanged. Using a data standard allows education stakeholders to have a baseline understanding of the quality of data (Austin et al. 2016). However, the language analogy used above is applicable in terms of data justice applications: the need to utilize one data standard results in exclusion of alternative data format or priority given to easily quantifiable and categorizable data. Data that is easily translatable into the data standard can be transmitted more easily, and data that is not easily translatable may get put on the technical priority back burner. An example of this exclusion is the narrative data held in the Individualized Education Program (IEP) required by the U.S. federal education law relating to students with disabilities. Data in the IEP is required to be individualized to the extent that student goals and progress toward those goals must be entered in narrative format only. The number of goals that a student may have varies, and the ways in which progress toward each goal is measured may vary as well.

Compare this to the measurement of educational progress for students without IEPs, which, in the U.S., is primarily measured by state standardized test scores, progress from grade level to grade level according to course grades or standards-based benchmarks, and local assessments given to all students in a school or school district. All of these measures have some form of quantitative or otherwise standardized score report (e.g., a letter), and are thus relatively easily transmitted via interoperable data systems. In implementations of interoperability, this results in less benefit to students with disabilities, as their primary education progress data points remain inaccessible, whereas general education students' primary data points become more accessible to instructors and leaders making decisions about needed interventions, funding priorities, and human resource allocations.

Although this example highlights the prioritization of easily transmittable data points, it also suggests that there may be opportunities for creative applications regarding the interoperability of data that is less standardized in its original format. For students with disabilities, goal creation and progress have often been relegated to a focus on policy compliance; however, recent U.S. legal precedent and guidance from the federal Department of Education have directed state and local education agencies to include a focus on measurable education progress results for students with disabilities (Endrew F. v. Douglas County School Dist. RE-1 2017; Office of Special Education Programs n.d.; Rowe et al. 2021). The technical tools available through the use of interoperability – especially in states where IEP data is housed in a system separate from the student information system – can provide school districts with more options to organize and analyze large amounts of IEP data alongside other data collected by schools. While IEP data is still in narrative format, interoperability can provide school districts with sets of data that include information on student engagement (attendance and behavior) as well as general indicators of learning (assessments and grades). Districts can use this information to understand whether subsets of students with disabilities (e.g., by specific disability, by race/ethnicity, by socioeconomic status; Abualghaib et al. 2019) are showing disproportionate outcomes on high-level measures, which can then be used to investigate further and more efficiently into the narrative data in the IEP. These data sets can also help district leaders to understand the quality of students' IEP goals and progress monitoring as entered by instructors, which can then help districts target training for educator improvement on these data pieces that are not only essential to the quality of education for students with disabilities, but that are required by federal and state law.

This example also illustrates the role of metadata (the context of data) in interoperability. Students' IEP goals – in the aggregate – can be examined according to when they were written, how often they are updated, and whether they are updated as a result of the student meeting the goal or due to the IEP team acknowledging that the current goal is inappropriate (too low or too ambitious), among other possibilities for goal analysis. Students' progress monitoring can be analyzed to see how often educators are updating progress information, what measures they are using, and how much progress is being made between updates, among other areas. However, this example also highlights how the analysis of metadata could be used for unjust purposes, such as to punish overburdened educators rather than support them.

Representation and categorization

Another potential area for data injustice that can occur with interoperability (though common in most if not all data scenarios involving state and/or federal requirements) is the limits imposed on categories and thus on individual identity and experience. Data standards can provide both opportunities and constraints for categories of identity validation and recognition in demographic categories that are socially constructed, such as gender identity and race/ethnic identity. For example, a data standard can allow local customization of allowable categories (such as non-binary or gender fluid identity), even if a state or federal entity does not permit such categories; this flexibility allows jurisdictions to provide gender-affirming options for student registration, which can significantly and positively impact student mental health (e.g., Conron and Austin 2008).

With race/ethnicity categories, the ability to capture all elements of a student's race/ethnic identity in the original, raw data then allows data analysts at all levels (district, state, federal, and external research/policy analysis) to customize the aggregation of categories for specific purposes. This customization is often needed in order to disaggregate and compare outcomes for student groupings which may be different across geographic location and time, which makes sense as these categories are constructed differently as society evolves. For example, U.S. federal categories include both race and 'ethnicity,' which is actually only a binary of 'Hispanic or non-Hispanic' (U.S. Census Bureau 2022).[1] According to the federal government, an individual can be Hispanic (or non-Hispanic) and be of any race. However, in some states, individuals identifying as Hispanic are stripped of their racial identity; data collection logic for race/ethnicity is 'if Hispanic, then "Hispanic"; if not Hispanic, but more than one race selected, then "2 or more races"; if not Hispanic and one race selected, then "selected race."'

Such logic may be useful to state data collections, but it introduces data justice questions about the authority of the state to decide which identity categories for an individual should be prioritized, and it also prevents all state-issued datasets from including more granular and accurate race/ethnicity identities of its constituents. The fact that they are issued by a governmental authority leads to a false assumption that those categories are objective and accurate representations of the identity of the individuals included within them (Fu and King 2021; Williamson and Piattoeva 2018). Without a raw, granular dataset, published data and associated analyses are already biased by high-level decisions of the state (Selwyn 2014). When state datasets are collected using interoperability, the raw datasets may provide a more accurate picture of the identities and experiences of a state's population. Related challenges to administrative datasets have been surfaced by previous researchers (e.g., Goldhaber, Holden, and Grout 2019). We recognize that the ability to disaggregate data by demographic and program categories can be both an opportunity for justice – highlighting disparate outcomes and providing the data justifying a change in practice (e.g., Abualghaib et al. 2019; Pang, Han, and Pang 2011) – as well as a risk of injustice, as it can have a negative impact both in how subgroups may view their own identities and/or how policymakers, teachers, and other stakeholders with power over students can transform disparate outcomes into biased judgments about subgroups (e.g., Copur-Gencturk et al. 2019; Schueler and West 2021; Scott et al. 2019; Smith et al. 2019).

Conclusion and future research

This paper has provided an overview of current and potential intersections of open-source interoperability initiatives on data justice for K-12 education communities. However, considering the lack of empirical research on interoperability and data standards in education (and in other social service/public sectors), this is certainly only a starting point. Much of the potential implications for data justice and injustice regarding interoperability are just that – potential – and research is needed to understand the extent to which these potential implications actually occur, the variables leading to and mitigating such occurrences, and the role of organizations at multiple levels and in multiple sectors (private and public) in those implications.

Based on our experiences with the design and implementation of interoperability initiatives, we recommend that policymakers and the public pay more attention to the *transparency* of these initiatives – how are the initiatives designed? What vendors have been contracted to implement or support implementation? How will local data be transformed into state data points? Will local agencies receive any benefit, when an interoperability initiative is required by a state agency? Currently, policies often written by state legislators are driving interoperability efforts, and these policies may not be aligned with technical requirements or best practices. Policymakers and technology vendors are primarily driving implementation choices in interoperability initiatives currently – not the needs of students, families, teachers, school leaders, and local communities. Like all areas of policy, interoperability initiatives can be designed with more or less transparency, quality, and justice in the details. For the public, learning more about what interoperability is as well as the potential benefits and risks is key to pushing for justice in these implementations.

Note

1. We recognize that the terminology 'Hispanic' is contested and problematic. However, we use the terms used by the jurisdictions of focus here in order to be consistent with the language of the jurisdictions' policy decisions.

Acknowledgments

The authors would like to thank Rosh Dhanawade (Education Analytics), Joel Boehner (South Bend Community School Corporation), and Matt Criscenzo (Indiana University) for insightful contributions to the content of this paper. The authors would also like to thank the anonymous reviewers for their incredibly helpful comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Molly S. Stewart http://orcid.org/0000-0002-8006-8798
Elizabeth Pier http://orcid.org/0000-0001-7309-5336

References

Abualghaib, O., N. Groce, N. Simeu, M. T. Carew, and D. Mont. 2019. "Making Visible the Invisible: Why Disability-Disaggregated Data is Vital to "Leave No-One Behind"." *Sustainability* 11: 3091. doi:10.3390/su11113091.

Allensworth, E. M., and J. Q. Easton. 2007. *What Matters for Staying on-Track and Graduating in Chicago Public High Schools: A Close Look at Course Grades, Failures, and Attendance in the Freshman Year*. Chicago, IL: University of Chicago Consortium on Chicago School Research.

Anagnostopoulos, D., and J. Bautista-Guerra. 2013. "Trust and Numbers: Constructing and Contesting Statewide Student Information Systems." In *The Infrastructure of Accountability: Data use and the Transformation of American Education*, edited by D. Anagnostpoulos, S. A. Rutledge, and R. Jacobsen, 41–56. Cambridge, MA: Harvard Education Press.

Anagnostopoulos, D., S. Rutledge, and V. Bali. 2013. "State Education Agencies, Information Systems, and the Expansion of State Power in the era of Test-Based Accountability." *Educational Policy* 27 (2): 217–247. doi:10.1177/0895904813475713.

Archibald, R. A. 2022. "The Rise of the Airport Metal Detector: Colorblind Racism, Police Discretion, and Surveillance Across Borders." *Journal of Social History* 56: 637–671. doi:10.1093/jsh/shac049.

Austin, C. C., T. Bloom, S. Dallmeier-Tiessen, V. K. Khodiyar, F. Murphy, A. Nurnberger, L. Raymond, et al. 2016. "Key Components of Data Publishing: Using Current Best Practices to Develop a Reference Model for Data Publishing." *International Journal on Digital Libraries* 18: 77–92. doi:10.1007/s00799-016-0178-2.

Balfanz, R., L. Herzog, and D. J. Mac Iver. 2007. "Preventing Student Disengagement and Keeping Students on the Graduation Path in Urban Middle-Grades Schools: Early Identification and Effective Interventions." *Educational Psychologist* 42 (4): 223–235. doi:10.1080/00461520701621079.

Chan Zuckerberg Initiative. 2023. Whole Child Approach to Education. https://chanzuckerberg.com/education/whole-child-approach-to-education/#:~:text=A%20whole%20child%20approach%20to,adults%20to%20engage%20and%20thrive.

Charteris, J. 2022. "Post-panoptic Accountability: Making Data Visible Through 'Data Walls' for Schooling Improvement." *British Journal of Sociology of Education* 43 (3): 333–348. doi:10.1080/01425692.2021.2018651.

Cho, V., and J. C. Wayman. 2014. "Districts' Efforts for Data use and Computer Data Systems: The Role of Sensemaking in System use and Implementation." *Teachers College Record* 116: 1–45. doi:10.1177/016146811411600203.

Cho, V., and J. C. Wayman. 2015. "Assumptions, Strategies, and Organization: Central Office Implementation of Computer Data Systems." *Journal of School Leadership* 25: 1203–1236. doi:10.1177/105268461502500607.

Clarke, R. 1988. "Information Technology and Dataveillance." *Communications of the ACM* 31 (5): 498–512. doi:10.1145/42411.42413.

Clutterbuck, J. 2022. "Data Infrastructures and the Governance of Their Accompanying Narratives." *British Journal of Sociology of Education* 43 (1): 120–139. doi:10.1080/01425692.2021.2003184.

Clutterbuck, J., I. Hardy, and S. Creagh. 2021. "Data Infrastructures as Sites of Preclusion and Omission: The Representation of Students and Schooling." *Journal of Education Policy* 38: 93–114. doi:10.1080/02680939.2021.1972166.

Conron, K. J., Scout, and S. B. Austin. 2008. "'Everyone has a Right to, Like, Check Their Box:' Findings on a Measure of Gender Identity from a Cognitive Testing Study with Adolescents." *Journal of LGBT Health Research* 4 (1): 1–9. doi:10.1080/15574090802412572.

Copur-Gencturk, Y., J. R. Cimpian, S. T. Lubienski, and I. Thacker. 2019. "Teachers' Bias Against the Mathematical Ability of Female, Black, and Hispanic Students." *Educational Researcher* 49 (1): 30–43. doi:10.3102/0013189X19890577.

Corbeil, M. E., J. R. Corbeil, and B. H. Khan. 2019. "A Framework for Implementing Responsible Data Mining and Analytics in Education." In *Responsible Analytics and Data Mining in Education: Global Perspectives on Quality, Support, and Decision Making*, edited by B. H. Khan, J. R. Corbeil, and M. E. Corbeil, 3–15. New York, NY: Routledge, Taylor & Francis Group.

Decuypere, M. 2021. "The Topologies of Data Practices: A Methodological Introduction." *Journal of New Approaches in Educational Research* 10 (1): 67–84. doi:10.7821/naer.2021.1.650.

Dencik, L., A. Hintz, and J. Cable. 2016. "Towards Data Justice? The Ambiguity of Anti-Surveillance Resistance in Political Activism." *Big Data & Society* 3 (2): 1–12. doi:10.1177/2053951716679678.

Dencik, L., A. Hintz, J. Redden, and E. Treré. 2019. "Exploring Data Justice: Conceptions, Applications and Directions." *Information, Communication & Society* 22 (7): 873–881. doi:10.1080/1369118X.2019.1606268.

Dexter, S., A. Francisco, and C. L. Luna. 2021. "Five Leading-Edge K-12 Districts' Decision-Making Processes for EdTech Innovations." *Journal of Educational Administration* 59 (3): 352–366. doi:10.1108/JEA-10-2020-0222.

Ed-Fi Alliance. 2023. https://www.ed-fi.org/.

Educating the Whole Child Summit. 2023. https://ismhi.indiana.edu/events/ishmi-whole-child-summit.html.

Education360. 2023. https://education360.com.au/.

Ehlers, L. 2020. Seeing the "Whole Child" Data Picture, June 17. https://www.illuminateed.com/blog/2020/06/seeing-the-whole-child-data-picture/.

Endrew F. v. Douglas County School Dist. RE-1, 580 U.S. ___. 2017.

Eubanks, V. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY: St. Martin's Press.

Eynon, R. 2013. "The Rise of Big Data: What Does it Mean for Education, Technology, and Media Research?" *Learning, Media and Technology* 38 (3): 237–240. doi:10.1080/17439884.2013.771783.

Fu, S., and K. King. 2021. "Data Disaggregation and its Discontents: Discourses of Civil Rights, Efficiency and Ethnic Registry." *Discourse: Studies in the Cultural Politics of Education* 42 (2): 199–214. doi:10.1080/01596306.2019.1602507.

Goldhaber, D., K. L. Holden, and C. Grout. 2019. "Errors in Administrative Education Data: A Cautionary Tale." *Educational Researcher* 48 (3): 179–182. doi:10.3102/0013189X19837598.

Gordon, W. J., and C. Catalini. 2018. "Blockchain Technology for Healthcare: Facilitating the Transition to Patient-Driven Interoperability." *Computational and Structural Biotechnology Journal* 16: 224–230. doi:10.1016/j.csbj.2018.06.003.

Gulson, K. N., and S. Sellar. 2019. "Emerging Data Infrastructures and the new Topologies of Education Policy." *Environment and Planning D: Society and Space* 37 (2): 350–366. doi:10.1177/0263775818813144.

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. doi:10.1177/1474904116648966.

Hawn Nelson, A., D. Jenkins, S. Zanti, M. Katz, E. Berkowitz, T. Burnett, and D. Culhane. 2020. *A Toolkit for Centering Racial Equity Throughout Data Integration*. Philadelphia, PA: Actionable Intelligence for Social Policy, University of Pennsylvania. https://www.aisp.upenn.edu/wp-content/uploads/2020/08/AISP-Toolkit_5.27.20.pdf.

Iwama, J., Y. Irizarry, A. Ernstes, M. Ripepi, A. A. Peguero, J. M. Bondy, and J. S. Hong. 2022. "Segregation, Securitization, and Bullying: Investigating the Connections Between Policing, Surveillance Punishment, and Violence." *Race and Justice* 1–32. doi:10.1177/21533687221105906.

Jaffe, R. 2020. "Rethinking Metadata's Value and how it is Evaluated." *Technical Services Quarterly* 37 (4): 432–443. doi:10.1080/07317131.2020.1810443.

Jarke, J., and A. Breiter. 2019. "Editorial: The Datafication of Education." *Learning, Media and Technology* 44 (1): 1–6. doi:10.1080/17439884.2019.1573833.

Karapetyan, A. 2018. Ensuring Student Success with 360-Degree Data Views, July 20. https://www.salesforce.org/blog/student-success-in-k-12-education-case-study-with-boston-day-and-evening-academy/.

Kerssens, N., and J. van Dijck. 2022. "Governed by Edtech? Valuing Pedagogical Autonomy in a Platform Society." *Harvard Educational Review* 92 (2): 284–303. doi:10.17763/1943-5045-92.2.284.

Kitchin, R. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. Los Angeles, CA: SAGE.

Komljenovic, J. 2021. “The Rise of Education Rentiers: Digital Platforms, Digital Data and Rents.” *Learning, Media and Technology* 46 (3): 320–332. doi:10.1080/17439884.2021.1891422.
Krachman, S. B., R. LaRocca, and C. Gabrieli. 2018. “Accounting for the Whole Child.” *Educational Leadership* 75 (5). https://www.ascd.org/el/articles/accounting-for-the-whole-child.
Landwehr, M., A. Borning, and V. Wulf. 2021. “Problems with Surveillance Capitalism and Possible Alternative for IT Infrastructure.” *Information, Communication & Society* 26: 70–85. doi:10.1080/1369118X.2021.2014548.
Lavis, T., and N. Brewer. 2017. “Effects of a Proven Error on Evaluations of Witness Testimony..” *Law and Human Behavior* 41 (3): 314–323. doi:10.1037/lhb0000217.
Leonelli, S. 2019. “Data — from Objects to Assets.” *Nature* 574: 317–320. doi:10.1038/d41586-019-03062-w.
Lindh, M., and J. Nolin. 2016. “Information we Collect: Surveillance and Privacy in the Implementation of Google Apps for Education.” *European Educational Research Journal* 15 (6): 644–663. doi:10.1177/1474904116654917.
Mandinach, E. B., and K. Schildkamp. 2021. “Misconceptions About Data-Based Decision Making in Education: An Exploration of the Literature.” *Studies in Educational Evaluation* 69. doi:10.1016/j.stueduc.2020.100842.
McKinney, E. H., and C. J. Yoos. 2010. “Information About Information: A Taxonomy of Views.” *MIS Quarterly* 34 (2): 329–344. doi:10.2307/20721430.
Murray, S. D., J. Hurley, and S. R. Ahmed. 2015. “Supporting the Whole Child Through Coordinated Policies, Processes, and Practices.” *Journal of School Health* 85: 795–801. doi:10.1111/josh.12306.
Noura, M., M. Atiquzzaman, and M. Gaedke. 2019. “Interoperability in Internet of Things: Taxonomies and Open Challenges.” *Mobile Networks and Applications* 24: 796–809. doi:10.1007/s11036-018-1089-9.
Office of Special Education Programs. n.d. *Slides to Explain Results Driven Accountability (RDA) [Part B]*. Washington, DC: Office of Special Education Programs, U.S. Department of Education. https://www2.ed.gov/about/offices/list/osers/osep/rda/part-b-rda-comprehensive-overview.ppt.
Pang, V. O., P. H. Han, and J. M. Pang. 2011. “Asian American and Pacific Islander Students: Equity and the Achievement Gap.” *Educational Researcher* 40 (8): 378–389. doi:10.3102/0013189X11424222.
Pang, C., and D. Szafron. 2014. “Single Source of Truth (SSOT) for Service Oriented Architecture (SOA).” *Proceedings of the 12th International Conference on Service Oriented Computing* 8831: 575–589. doi:10.1007/978-3-662-45391-9_50.
Pangrazio, L., N. Selwyn, and B. Cumbo. 2022. “A Patchwork of Platforms: Mapping Data Infrastructures in Schools.” *Learning, Media and Technology* 48: 65–80. doi:10.1080/17439884.2022.2035395.
Perrotta, C., K. N. Gulson, B. Williamson, and K. Witzenberger. 2021. “Automation, APIs and the Distributed Labour of Platform Pedagogies in Google Classroom.” *Critical Studies in Education* 62 (1): 97–113. doi:10.1080/17508487.2020.1855597.
Phelps, D., and R. Santo. 2022. “Debugging Inequities: Data use, “Gumshoe Work,” and Problem Identification in District-Wide Computer Science Education Initiatives.” *Policy Futures in Education* 1–23. doi:10.1177/14782103221126317.
Piety, P. J. 2019. “Components, Infrastructures, and Capacity: The Quest for the Impact of Actionable Data use on P-20 Educator Practice.” *Review of Research in Education* 43: 394–421. doi:10.3102/0091732X18821116.
Rafalow, M. H., and C. Puckett. 2022. “Sorting Machines: Digital Technology and Categorical Inequality in Education.” *Educational Researcher* 51 (4): 274–278. doi:10.3102/0013189X211070812.
Ralyea, D., and A. Rice. 2022. Data Neighborhoods: Preserving Context Around Educational Data, April 4. https://www.edanalytics.org/blog/data-neighborhoods-preserving-context-around-educational-data.
Rowe, D. A., C. H. Fowler, C. D’Agord, F. Horiuchi, M. Kawatachi, G. C. Norbert, and S. K. Avoke. 2021. “State Systemic Improvement Planning: Impact on System and Student Outcomes.” *Journal of Disability Policy Studies* 32 (2): 131–141. doi:10.1177/1044207320932548.
Schueler, B. E., and M. R. West. 2021. “Federalism, Race, and the Politics of Turnaround: U.S. Public Opinion on Improving low-Performing Schools and Districts.” *Educational Researcher* 51 (2): 122–133. doi:10.3102/0013189X211053317.
Scott, T. M., N. Gage, R. Hirn, and H. Han. 2019. “Teacher and Student Race as a Predictor for Negative Feedback During Instruction.” *School Psychology* 34 (1): 22–31. doi:10.1037/spq0000251.
Sellar, S. 2015. “Data Infrastructure: A Review of Expanding Accountability Systems and Large-Scale Assessments in Education.” *Discourse: Studies in the Cultural Politics of Education* 36 (5): 765–777. London, UK: Bloomsbury Academic. doi:10.1080/01596306.2014.931117.
Sellar, S. 2017. “Making Network Markets in Education: The Development of Data Infrastructure in Australian Schooling.” *Globalisation, Societies and Education* 15 (3): 341–351. doi:10.1080/14767724.2017.1330137.
Sellar, S., and K. N. Gulson. 2019. “Dispositions and Situations of Education Governance: The Example of Data Infrastructure in Australian Schooling.” In *Education Governance and Social Theory: Interdisciplinary Approaches to Research*, edited by A. Wilkins, and A. Olmedo, 63–79. New York, NY: Bloomsbury Academic.
Selwyn, N. 2014. “Data Entry: Towards the Critical Study of Digital Data and Education.” *Learning, Media and Technology* 40 (1): 64–82. doi:10.1080/17439884.2014.921628.
Selwyn, N. 2016. “‘There’s so Much Data’: Exploring the Realities of Data-Based School Governance.” *European Educational Research Journal* 15 (1): 54–68. doi:10.1177/1474904115602909.

Selwyn, N. 2021. "The Human Labour of School Data: Exploring the Production of Digital Data in Schools." *Oxford Review of Education* 47 (3): 353–368. doi:10.1080/03054985.2020.1835628.

Selwyn, N. 2022. "'Just Playing Around with Excel and Pivot Tables' – The Realities of Data-Driven Schooling." *Research Papers in Education* 37 (1): 95–114. doi:10.1080/02671522.2020.1812107.

Selwyn, N., L. Pangrazio, and B. Cumbo. 2022. "Knowing the (Datafied) Student: The Production of the Student Subject Through School Data." *British Journal of Educational Studies* 70 (3): 345–361. doi:10.1080/00071005.2021.1925085.

Smith, P., A. Kumi-Yeboah, R. Chang, J. Lee, and P. Frazier. 2019. "Rethinking "(Under) Performance" for Black English Speakers: Beyond Achievement to Opportunity." *Journal of Black Studies* 50 (6): 528–554. doi:10.1177/0021934719851870.

State Educational Technology Directors Association. 2018. State Education Leadership Interoperability: Leveraging Data for Academic Excellence (ED590233) *ERIC*. http://files.eric.ed.gov/fulltext/ED590233.pdf.

Taylor, L. 2017. "What is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally." *Big Data & Society* 4: 1–14. doi:10.1177/2053951717736335.

U.S. Census Bureau. 2022. About the Topic of Race. https://www.census.gov/topics/population/race/about.html.

Weatherly, R., and M. Lipsky. 1977. "Street-level Bureaucrats and Institutional Innovation: Implementing Special-Education Reform." *Harvard Educational Review* 47 (2): 171–197. doi:10.17763/haer.47.2.v870r1v16786270x.

Whole Child Conference. 2023. https://www.njascd.org/store/p/5th-annual-whole-child-conference.

Whole Child Summit. 2023. https://www.characterstrong.com/wholechildsummit2023/.

Wikman, C., M. W. Allodi, and L. A. Ferrer-Wreder. 2022. "Self-concept, Prosocial School Behaviors, Well-Being, and Academic Skills in Elementary School Students: A Whole-Child Perspective." *Education Sciences* 12: 298–315. doi:10.3390/educsci12050298.

Williamson, B. 2015. "Governing Software: Networks, Databases and Algorithmic Power in the Digital Governance of Public Education." *Learning, Media and Technology* 40 (1): 83–105. doi:10.1080/17439884.2014.924527.

Williamson, B. 2016. "Digital Education Governance: Data Visualization, Predictive Analytics, and 'Real-Time' Policy Instruments." *Journal of Education Policy* 31 (2): 123–141. doi:10.1080/02680939.2015.1035758.

Williamson, B. 2022. "Big Edtech." *Learning, Media and Technology* 47 (2): 157–162. doi:10.1080/17439884.2022.2063888.

Williamson, B., K. N. Gulson, C. Perrotta, and K. Witzenberger. 2022. "Amazon and the new Global Connective Architectures of Education Governance." *Harvard Educational Review* 92 (2): 231–256. doi:10.17763/1943-5045-92.2.231.

Williamson, B., and N. Piattoeva. 2018. "Objectivity as Standardization in Data-Scientific Education Policy, Technology and Governance." *Learning, Media and Technology* 44 (1): 64–76. doi:10.1080/17439884.2018.1556215.