

The promise and perils of education data

Education budgets are tied to how well a school performs, as demonstrated by data. Education data interoperability is called for to maintain institutional legitimacy and bring educational benefits. In [this US study](#), Dr Velislava Hillman asks, what are the perils of education data for children?

In 2012, the Bill & Melinda Gates Foundation funded [InBloom](#) – a single warehouse that would collect a massive store of confidential and sensitive student and teacher data (Figure 1). This resulted in several years of parental uproar by parents fearful for children’s privacy that finally [brought it down](#).

Figure 1: According to InBloom’s objectives, teachers would use data-generated dashboards to monitor students’ progress at academic and non-academic levels.

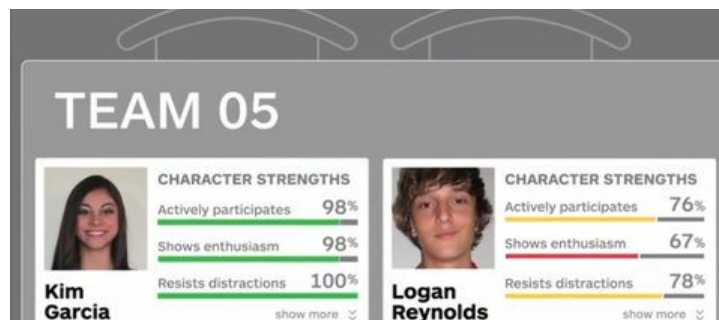


Image credit: HBS Digital Initiative

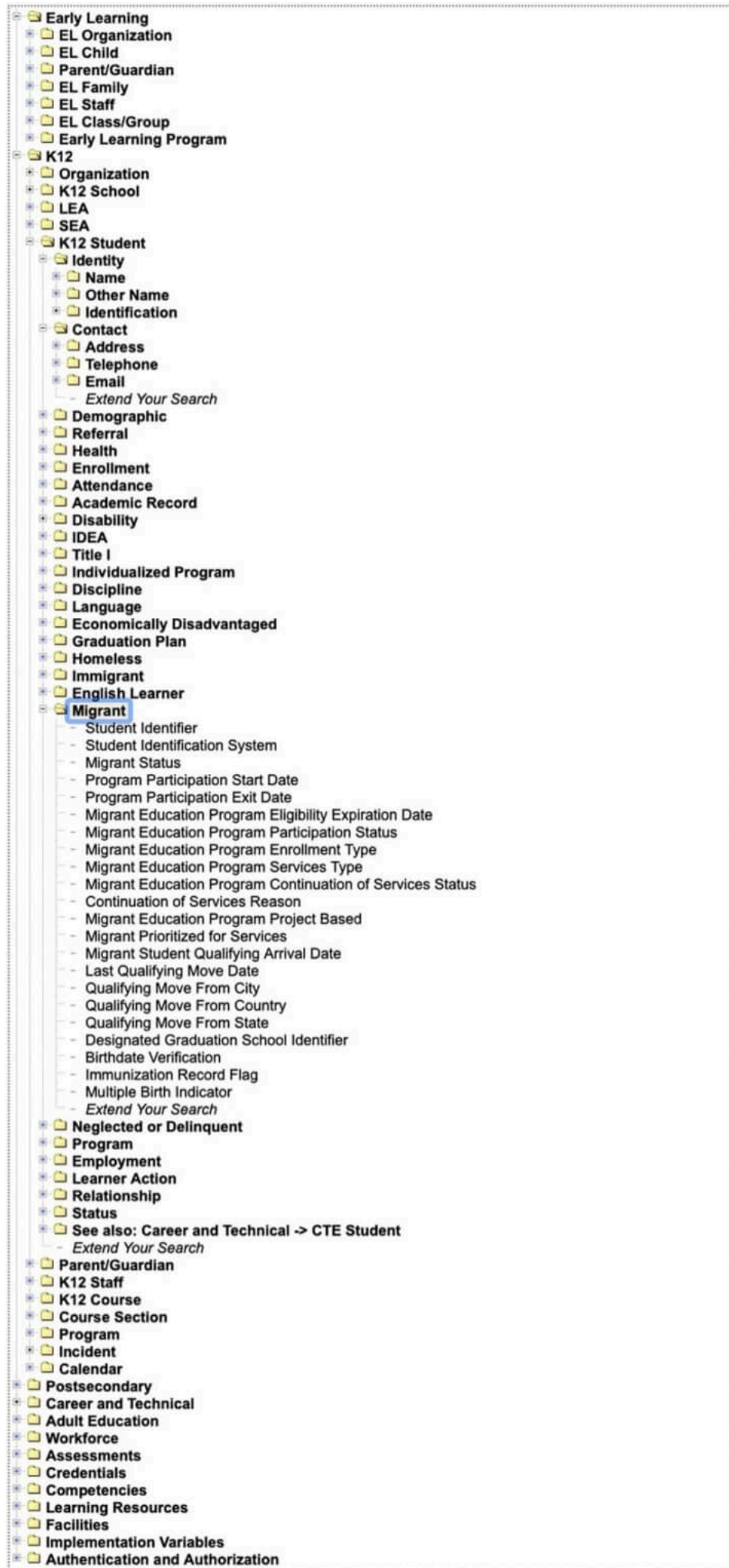


Figure 2. Examples of CEDS's data elements ,

Ten years later, little has changed. We see massive data system assemblages and education technology (EdTech) solutions offering interoperable and granular data generation and concentration, together with advanced analytics and student profiling with capacities for behavioural modification and predetermined futures.

The promise of data

In the classroom, teachers refer to data for the purpose of instruction, pedagogical intervention and progress monitoring. Many EdTech applications provide digital dashboards, data visualisations, analytics and recommendations. At district and country levels (in US and UK), data systems enable educational institutions to provide accountability reports.

However, as EdTech platforms and apps diversify in type and increase in use, so do data which are not always compatible with student management systems. Data interoperability necessitates common standards of how data are collected and organised. In a nutshell, some systems' data may tell one thing about a student, but putting together all systems that operate – will (supposedly) give a fuller picture.

In the US, the Common Education Data Standard (CEDS) is a data interoperability template available for districts and states to adopt. Reminiscent of InBloom, CEDS's Warehouse is an 'initiative ... to streamline the understanding between and across P-20W' – data across the early years, kindergarten, primary, postsecondary and workforce institutions. CEDS's proposed 'common language' contains hundreds of data elements (Figure 2).

The perils of education data

1. The introduction of a common language of data raises concerns. A common data language streamlines what information needs collecting and how to organise it and, thereby, influences what counts in education and how to count it. Such language serves schools well in providing their accountability reports. But does it serve a child's interest to learn and in what way? This question requires a pedagogical and curriculum expertise to critically evaluate how data may be impacting or even transforming education since data impose their own 'grammars of action' and can 'oversimplify the activities they are intended to represent'.

2. The language of data interlocks an expanding network of third parties that adopt it. The CEDS Data Warehouse has the capacity to support the full P-20W data pipeline (Figure 3). It partners with learning agencies, public and private higher education institutions, the US Department of Education, the US Health and Human Services and the US Department of Labour, and others. Note that while InBloom failed, the Gates Foundation's two other initiatives striving for 'pathways data' – the Data Quality Campaign and Chiefs for Change, a bipartisan lobby group push aggressively for data alignment and student tracking while also partnering with CEDS.

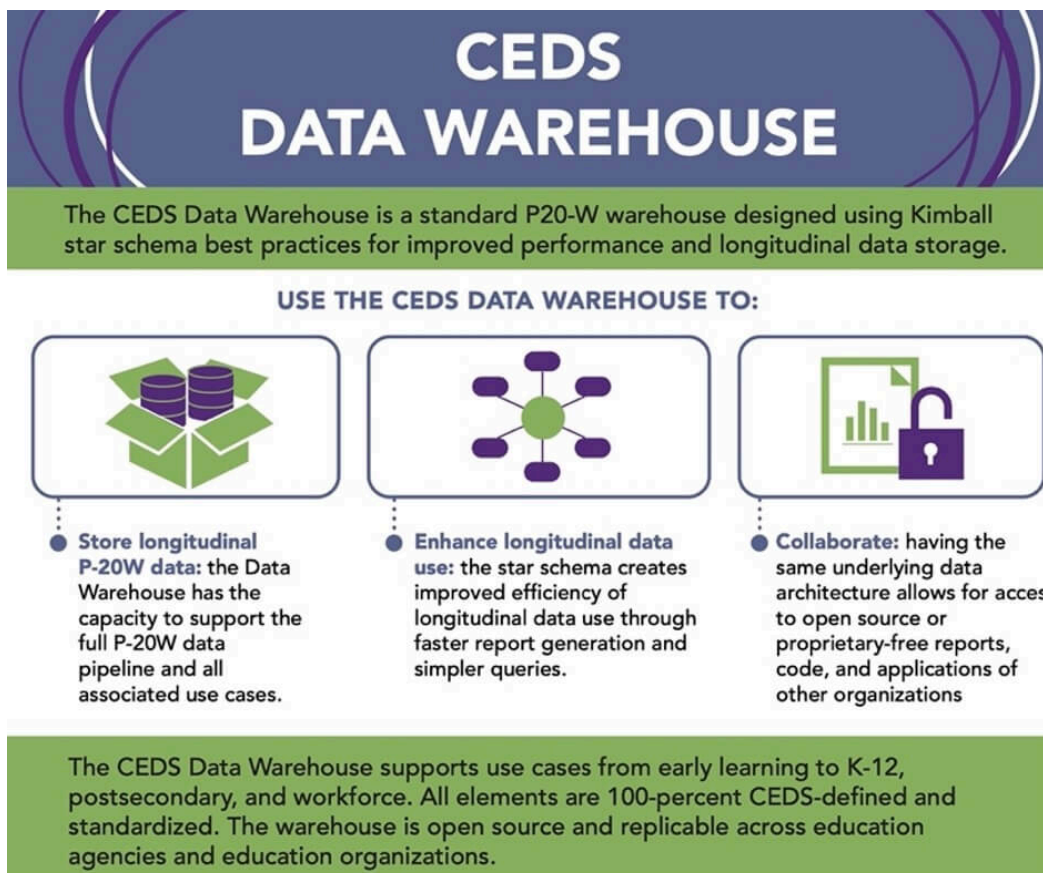


Figure 3: CEDS Data Warehouse envisioning a pool of data that drives the cradle-to-career pipeline development which powerful corporations, such as Amazon, embrace.

3. The language of data for data interoperability has led to data pipeline development across districts and states and a new way of thinking about education. States are developing *datamarts* – dashboards – for access to ‘targeted sets of data related to specific topics or questions’. Others are introducing data ‘lakes’, an euphemism for a warehouse, ‘capable of ingesting, storing, and providing data from a large number of sources and for a wide range of users and uses’. Yet others have been promoting data ‘backpacks’ or electronic student records that contain all sorts of academic and personal student data – test scores, behavioural patterns, ‘non-cognitive variables that impact achievements, as well as an ‘early warning system’, self-management skills, behaviour/character education, and a record of community service’.

In other words, the language of data and the capacities stemming from data interoperability have enabled for a direct link to be made between education and the labour market. For example, aligning education with industry through the common language of data has led to clustering Washington and Virginia into industrial zones. Education policies follow by proposing hyper-specialised education and training depending on which zone a student comes from, leaving parents and students concerned.

In the UK, learning-to-earning models that align education with industry demands are also on the way. The European Union, too, has agendas for harvesting skills intelligence through the development of a ‘permanent online tool for real-time information’ for ‘all interested stakeholders’ to tailor careers and inform education policy based on industry demands. In a word, a common data-driven language allows for policy and industry to ‘speak’ about what needs training depending on what labour is in demand.

This model suggests that education data serve mainly industry and its labour demands, which may not necessarily advance the best interests of young learners. This further begs to question whether data pipelines can reduce education into the sole means for labour production – and even class (re)production.

4. The language of data has the capacity to modify the conditions which shape a learner's behaviour. The learning process is fitted to a formulaic if-then pattern such that *what* is taught, *how* it is taught and what is *learned* is quantified through numbers – data. This assumes that education data will benefit the system – be that of education or industry – on which the student is a reactive, modifiable participant (a node).

If data are used to shape children's learning opportunities, such use of data could limit the agency of young learners and interfere with their rights to education.

Meaningful mechanisms of oversight must be set up, based on a clear assessment of what education is for, and what children – and society – need and expect from it to diminish the perils and benefit from data. Only then can we consider what data collection will fulfil these expectations.

Note: This paper is available upon request in PDF format from the author.

Velislava Hillman is a Visiting Fellow at LSE and founder of EDDS, where she leads an independent team of international experts providing comprehensive audit and evaluation of education technology operators. As a researcher and academic Dr Hillman's work lies in education focused on the integration of AI systems into schools and the role and participation of children and young people in increasingly digitalised learning environments. Through EDDS, Dr Hillman maintains two objectives: to develop minimum standards and benchmarking of the edtech sector, and to offer a unique enhanced reporting mechanism that gives students, parents and educators peace of mind when choosing and using education technologies.



Originally posted on <https://digitalfuturescommission.org.uk/> on May 9, 2022 .