Working Group Report on Data for Learning

The Transformative Potential of Data for Learning

September 2023

BROADBAND COMMISSION
FOR SUSTAINABLE DEVELOPMENT

ITU

unesco

Working Group Report on Data for Learning

The Transformative Potential of Data for Learning

September 2023

unesco

Acknowledgements

This report was written collaboratively, with members of the Broadband Commission for Sustainable Development's Working Group on Data for Learning contributing their knowledge and expertise. The Working Group was chaired by Audrey Azoulay (Director-General, UNESCO) and coordinated by Borhene Chakroun, Director of the Division of Policies and Lifelong Learning Systems, UNESCO, who guided the production of this report.

The UNESCO writing team included Juliette Norrmén-Smith and Saurabh Roy, with contributions from Carmela Salzano, Patrick John Devaney and Wayne Holmes. The report would not have been possible without the guidance of the Broadband Commission Secretariat team, including Anna Polomska, Leah Mann and Julia Gorlovetskaya.

The views expressed here are not attributed to any one organization or individual, except in the case studies contributed by respective Working Group members, nor do the views necessarily reflect the position of the Broadband Commission members or their affiliated organizations. Not all Working Group members participated in every meeting that led to the content of this report. We wish to thank the Working Group's Broadband Commissioners, commissioners' focal point members and external experts for their invaluable contributions, kind reviews and helpful comments.

We are grateful to the team from Strategic Agenda, including Liyana Aini, Gabriel Polley and Lauren Harper, Darcy Bush, and Julia Spaeth and Robert Tubb, for the editing, design and layout of this report, and the accompanying job board. The production of both documents was overseen by Juliette Norrmén-Smith.

Disclaimer

Working Group members

Commissioners and focal points

- Dr Tedros Ghebreyesus, World Health Organization (represented by Mr Derrick Muneene)
- Dr Carlos M Jarque, America Movil
- Baroness Beeban Kidron, 5Rights Foundation (represented by Ms Sonia Livingstone)
- Mr Lacina Koné, Smart Africa (represented by Ms Aretha Mare, Ms Thelma Efua Quaye, Mr Calvain Nangue and Mr Osman Issah)
- Professor Jeffrey Sachs, The Earth Institute, Columbia University (represented by Dr Radhika Iyengar)
- Professor Mercedes Aráoz, Peru
- Dr Lee Yee Cheong, Former Broadband Commissioner, International Science Technology and Innovation Centre for South-South Cooperation, UNESCO
- Mr Amir Dossal, Global Partnerships Forum
- Mr Bocar Ba, SAMENA Telecommunications Council
- Mr Erik Ekudden, Ericsson (represented by Ms Zohra Yermeche)
- Mr Achim Steiner, United Nations Development Programme (represented by Mr Gayan Peiris)
- Dr Rumman Chowdhury, Parity Responsible Innovation Fund
- Dr Qu Dongyu, Food and Agriculture Organization of the United Nations (represented by Mr Maximo Torero, Mr José Rosero Moncayo and Mr Henry Burgsteden)
- Mr Mauricio Ramos, Millicom
- Mr Piotr Dmochowski-Lipski, EUTELSAT IGO
- Ms Sima Sami Bahous, UN Women

External experts

- Ms Maria Florencia Ripani, Fundación Ceibal
- Mr Dale Allen, DXtera Institute/EdSAFE AI Alliance
- Mr Alexandre Barbosa, Cetic.br, UNESCO Regional Centre for Studies on the Development of the Information Society
- Mr Fabio Senne, Cetic.br, UNESCO Regional Centre for Studies on the Development of the Information Society
- Mr Werner Westermann Juárez, Library of National Congress of Chile
- Ms Carmela Salzano, Global Partnership for Education
- Mr Amit Prakash, Centre for IT and Public Policy, International Institute of Information Technology Bangalore
- Mr Wayne Holmes, Institute of Education, University College London
- Mr Doug Casey, Connecticut Commission for Educational Technology
- Ms Karine Sonigo, International Labour Organization
- Ms Beth Havinga, European EdTech Alliance
- Mr Ashok Goel, National AI Institute for Adult Learning and Online Education
- Dr Joanna Rubinstein, former Broadband Commissioner and Council Member of the UNSDSN
- Mr Shiloh Naiken, National Education Collaboration Trust, South Africa
- Ms Tracey Burns, National Center on Education and the Economy
- Ms Pascale Raulin-Serrier, Commission Nationale de l'Informatique et des Libertés
- Mr Xavier Delporte, Commission Nationale de l'Informatique et des Libertés
- Mr Jun Yu, National University of Singapore
- Ms Velislava Hillman, London School of Economics
- Mr Stéphan Vincent-Lancrin, OECD
- Ms Carol O'Donnell, Smithsonian Science Education Center
- Ms Kati Tiainen, Microsoft
- Ms Ellen J. Helsper, Department of Media and Communications, London School of Economics
- Ms Lidija Kralj, European Union
- Mr Shitanshu Mishra, UNESCO Mahatma Gandhi Institute of Education for Peace and Sustainable Development

List of tables, figures and boxes

Tables

Figures

Boxes

Acronyms and abbreviations

AI	Artificial intelligence
API(s)	Application programming interface(s)
Cetic.br	Regional Center for Studies on the Development of the Information Society
CNIL	Commission Nationale de l'Informatique et des Libertés
CTE	Career Technical Education
EdTech	Educational technology
EER	Education Ecosystem Registry
EMIS	Education management information system
GDPR	General Data Protection Regulation
ICT(s)	Information and communication technology(ies)
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais
IoT	Internet of Things
IPR	Intellectual property rights
IT	Information technology
ITU	International Telecommunication Union
LLM	Large language model
LMIC(s)	Low- and middle-income country(ies)
LMS	Learning management system
MIS	Management information system
NGO	Non-governmental organization
NIC.br	Brazilian Network Information Centre
OECD	Organisation for Economic Co-operation and Development
PPP(s)	Public-private partnership(s)
SA-SAMS	South African School Administration and Management System
SDG(s)	Sustainable Development Goal(s)
SIGED	Sistema de Información y Gestión Educativa
TVET	Technical and vocational education and training
UDISE+	Unified District Information System for Education Plus
UIS	UNESCO Institute for Statistics
UNCTAD	United Nations Conference on Trade and Development
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations Children's Fund
USAF	Universal service and access funds
VICT	Computational Thinking for Persons with Visual Impairment
VSK	Vidya Samiksha Kendra
WGD4L	Working Group on Data for Learning

Contents

UNESCO foreword

By 2025, it is projected that the world will generate a dizzying 175 trillion gigabytes of data per year – 88 times more than in 2010. If harnessed properly, these data could create new opportunities in the education field, improving teaching and learning processes, as well as the management and administration of educational institutions.

However, data are not a magic bullet, capable of improving education for everyone worldwide. A number of asymmetries affect data for learning across the globe. These gaps – in information, skills and sovereignty – are described in this report, which has been produced by the Broadband Commission Working Group on Data for Learning, co-chaired by UNESCO.

Firstly, regarding information, our report underlines that 60 per cent of primary schools, 50 per cent of lower secondary schools and 35 per cent of upper secondary schools are still not connected to the Internet. This means that learning data fail to capture disconnected learners, who often reside in rural, remote and disadvantaged locations.

Secondly, the report highlights a gap in skills. Not only do some teachers and learners lack the competencies to navigate the digital world – unsurprising given that 46 per cent of countries have not established digital skills standards in education. There is also insufficient understanding of the risks that data collection poses for human rights, at a time when only 16 per cent of countries have passed laws guaranteeing data privacy in education.

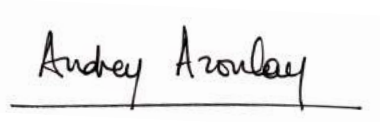
This leads to a third major issue: asymmetries in sovereignty. Nowadays, the governance of data is increasingly concentrated in the hands of a privileged few, in particular in the private sector, with implications for the rights of learners and teachers. Our report therefore argues for the creation of ethical standards for data collection, use and sharing across borders.

Harnessing the power of international cooperation, UNESCO has developed crucial building blocks to steer data in education in ethical ways, through our work with the Broadband Commission, of course, but also through our Recommendations on the Ethics of Artificial Intelligence and Open Science, adopted in 2021. These normative instruments provide countries with a framework allowing them to adapt their legislation, policies and programmes.

Our actions in this field reflect our belief that educational policies and practices may be data-informed, but not data-driven. In other words, data collection must never be an end in itself; it should always serve a purpose. This is the conclusion of this publication, which also resonates with the recently released UNESCO Global Education Monitoring Report, "Technologies in Education: A Tool on Whose Terms?"

Now, together, we must commit to the pathways set out in this document, to ensure that data for learning truly serves the public interest. And we must do this now, because new technological disruptions are waiting in the wings, from generative AI to neurotechnology.

In all these evolutions, UNESCO believes that one conviction must act as our compass: the idea that technology must serve learners first, not the other way around. We must steer the data revolution in education on our own terms.

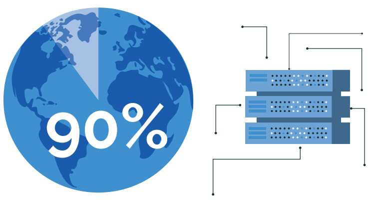


Audrey Azoulay
Director-General of UNESCO

Executive summary

Data for learning: A double-edged sword

While education data has long been collected, recent advances in data-fuelled digital technologies and global telecommunications networks have significantly amplified the volume and complexity of learning data flows. Globally, data production is growing exponentially, with 90 per cent of the world's digital data generated in the last two years alone. Recent estimates predict that the world will generate **175 trillion** gigabytes of data per year by 2025, which is almost 88 times greater than the amount produced in 2010.



of the world's data was generated in the last two years alone
(Statista, 2022)

Despite this global upward trend, data in the education sector is only valuable if it is collected and shared in a timely, transparent and trustworthy manner, and used for the specific purpose of improving learning, teaching, administration or strategic management.

Today, education and training systems around the world are struggling to deploy their agency to steer the data revolution on their own terms, towards targeting persistent education challenges and strengthening system readiness for a rapidly changing world.

This is due, in part, to deep digital divides within and across countries. Globally, **one in three people** do not use the Internet, blocked by lack of access, affordable data, or quality connection. Despite significant international efforts to connect every school to the internet, as exemplified by the progress made by the UNICEF-ITU Giga initiative, universal school connectivity remains limited, as **25 per cent** of primary schools worldwide do not have electricity. As a result, digital education data do not flow from many rural and developing areas, creating stark inequalities in the ability to use learning data to generate insights to improve the quality of teaching and learning and include the most marginalized.

These systemic obstacles are significant. With an estimated **773 million** illiterate adults and **244 million** young people out school – in addition to the hundreds of millions of people in the workforce who require retraining – all levels of the education workforce need data competencies combined with access to quality data to: make informed decisions on how best to manage policies, budgets, resources and classrooms; adapt to the digital transformation occurring in all sectors; and develop innovative solutions to counter the global learning crisis.

There is a clear need to build the capacities of the education and training workforce to have the agency to steer the data revolution in the service of global educational equity, quality and inclusion. Supporting the growth of the sector's holistic data literacies – beyond technical expertise to include competences related to governance, regulation, compliance and social impact – requires strengthening alliances with the private sector, civil society and research institutions.

> **"If harnessed properly, the digital revolution could be one of the most powerful tools for ensuring quality education for all and transforming the way teachers teach and learners learn. But if not, it could exacerbate inequalities and undermine learning outcomes."**
>
> *United Nations Secretary-General António Guterres at the Transforming Education Summit, 2022*

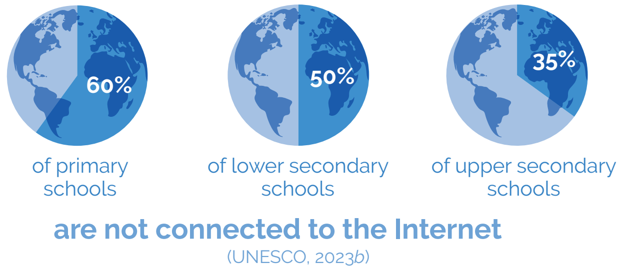
In September 2022, the United Nations convened the **Transforming Education Summit**, reigniting a collective commitment to lifelong learning as a pre-eminent public good, with the urgent goal of transforming education to be relevant and responsive in the digital era. The Secretary-General of the United Nations cautioned that the digital revolution could exacerbate inequalities and undermine learning outcomes if the international community did not urgently come together to overcome the digital divide and reinforce the capacity of education and training systems to steer the process and promote its agency including through expertise, capabilities and institutions, as well as high-quality contextualized digital learning content, backed by strong, sovereign and secure data systems.

Designing, implementing, governing and monitoring such information systems is no simple task in the complex education sector, the many asymmetries of which impede synergized approaches to data for learning.

Asymmetry of information: Swift and sweeping social changes vs slow and siloed data flows

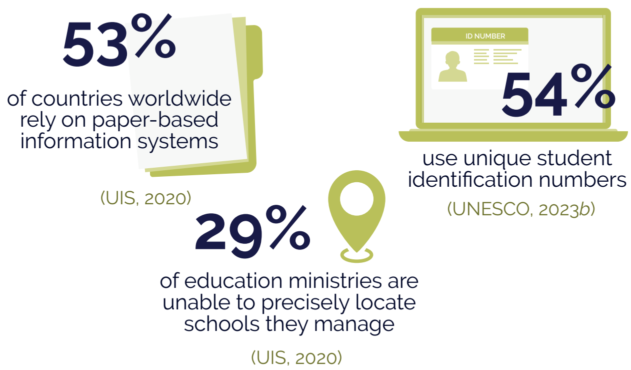
Education systems face growing pressures to rapidly respond to the many transitions unfolding around the world, from demographic and labour-market shifts, to numerous digital revolutions, to climate change. However, the sector does not always have – or use – the available information to anticipate the impact of these changes. As a result, education management often becomes a reactive rather than proactive process, unprepared to mitigate crises.

Much of the world does not have access to timely, quality and complete education data. Globally, **60 per cent** of primary schools, **50 per cent** of lower secondary schools and **35 per cent** of upper secondary schools are not connected to the Internet (UNESCO, 2023*b*). These data make clear that the effective and equitable use of digital data for learning is deeply intertwined with the digital divide. Disconnected learners, as well as marginalized learners and women and girls, risk being under-represented in education sector datasets. The invisibility of certain learners and disproportionate representation of others in these datasets may result in the further marginalization of disconnected communities whose needs remain unseen by policy-makers.



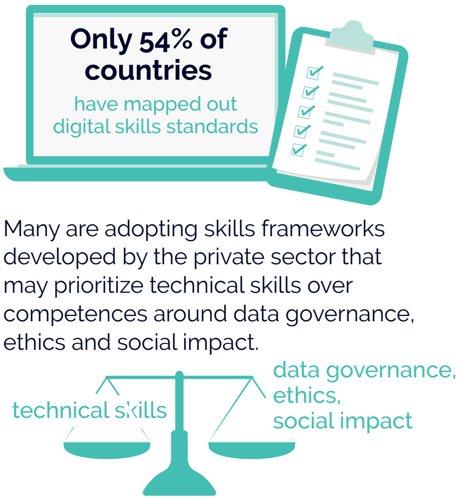
While the right to education is increasingly recognized as lifelong, beginning at birth and continuing through life, data collected in different learning settings – from schools, to workplaces, to community centres – are often disconnected, non-interoperable or lacking comparable standards. As a result, an individual's accomplishments or competencies may not be recognized or transferable, thus impeding their access to the labour market or further learning opportunities. This gets further exacerbated in the case of learners who are displaced due to forced or unforced migrations, and often lack physical records.

If information flows were better coordinated and connected both vertically across all levels of the education sector, and horizontally with the data collected by other sectors, such as health, finance, communication and labour-market data, education leaders and institutions could strengthen the resilience and resonance of their policies and practices. National, regional and global information observatories, such as those on labour-market changes, population shifts, climate risks, evolving technologies or Sustainable Development Goal (SDG) benchmarking, can be valuable analytic and comparative tools for the education sector to introduce into regular policy-making and planning cycles.



Asymmetry of skills: Big demands for "data-*driven*" vs little literacy for "data-*informed*"

There is growing pressure to use data in decision-making at every level of the education system. Therefore, every actor in the education system should have the necessary data-related competences to ensure effectiveness, accountability, compliance, privacy and security in the use of education data to improve quality and equity in all learning settings. While *data-driven* decisions tend to be based on a direct surface reading of data, *data-informed* approaches are inferred from a contextualized and critical interpretation of education data – one that balances the data with human insights, and shared principles of ethical and responsible use.

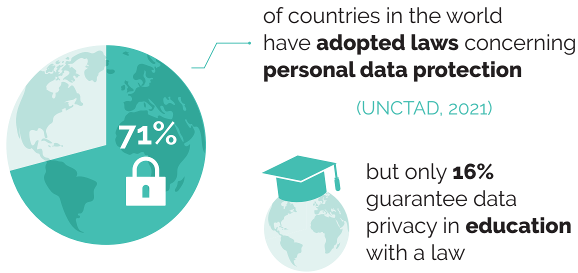


(UNESCO, 2023*b*)

To date, only **54 per cent** of countries have established digital skills standards, and many are adopting skills frameworks developed by the private sector that may prioritize technical skills that are valuable in commercial settings (UNESCO, 2023*b*). However, digital data literacy is not limited to technical or software expertise, such as a strong understanding of data analytics tools, cybersecurity practices, and emerging technologies such as artificial intelligence (AI) and virtual reality (VR). It is also essential to be well-versed in the social implications of data use in education, and in particular, the challenges it raises regarding inclusion, equity, ethics, ownership, agency of teachers and learners, and environmental and financial sustainability. Critically assessing the benefits and risks of data use in every learning experience is a key component of data literacy, which must be grounded in an understanding of what data represents, what it does not represent, and, indeed, what it may misrepresent.

Both critical data skills and technical digital competences are lacking around the world, posing significant barriers to the safe and effective use of data in education. Given algorithmic technologies fuelled by large amounts of data, such as applications of AI, are increasingly integrated into education systems, it is necessary that all education stakeholders are able to communicate data insights, assess data quality, grasp the main principles of data governance and ownership, and understand the impact of data use on people and human rights. Investments in capacity-building would reinforce the agency of education and training institutions to direct, design and drive the data revolution towards improving learning opportunities for all.

Asymmetry of sovereignty: Local legal frameworks vs global data flows



(UNESCO, 2023*b*)

Data in education are not static. They are collected, processed, transferred, stored, combined, separated, archived and/or destroyed, often in different settings. These settings may not be the same classroom, school district, country or even continent. A growing concentration of data is in the hands of a few large corporations and countries located in the global North. Without adequate legislation or public awareness, the current imbalance in data power could restrict user autonomy over their data and compromise national data sovereignty and security. To avoid undermining education as a universal human right, it is crucial to assess how their support can benefit resource-scarce education systems.

> "Despite the desire to make education a global common good, the role of commercial and private interests in education continues to grow, with all the ambiguities that entails: to date, only one in seven countries legally guarantees the privacy of educational data."
>
> *Audrey Azoulay, Director-General of UNESCO*

Under the right conditions, cross-border data flows within a lifelong learning perspective could expand the scope of educational opportunities and foster more culturally diverse and inclusive learning environments. For example, students and educators from different backgrounds could connect and access a broad range of educational resources from various countries, including online courses, digital libraries and learning materials. Global standards for safe and secure education data sharing could also improve learner and worker international mobility, paving the way for improved mechanisms for the cross-border recognition of learning outcomes.

> Governments should adopt and implement legislation, standards and agreed good practices to protect learners' and teachers' human rights, well-being and online safety, taking into account screen and connection time, privacy, and data protection; to ensure that data generated in the course of digital learning and beyond are analysed only as a public good; to prevent student and teacher surveillance; to guard against commercial advertising in educational settings; and to regulate the ethical use of artificial intelligence in education.
>
> *(UNESCO, 2023b)*

For these potential benefits to be unlocked, multilateral cooperation and solidarity are necessary to navigate complexities such as intellectual property legislation, licencing and cybersecurity. Despite the global nature of data flows, minimal international cooperation on data governance and regulation impedes the realization of data for learning as a common, public good. Cross-border, multilateral normative instruments can establish ethical principles for cooperation between public and private institutions, for these partnerships can spur innovation and bolster operational capacities to develop and deliver targeted management and learning tools designed to improve education ecosystems. However, given the role education data may play in business plans, it is the international community's collective responsibility to ensure that access to information on learning and education more broadly is not only reserved for those who can pay for it, and that the knowledge commons remains a shared, community-owned resource.

In the face of these asymmetries, three essential questions guide this report on the transformative potential of data for learning:

1. How can we maximize the benefits and minimize the risks of data use in education?
2. How can we ensure data is used to improve quality and equity in all learning settings?
3. How can we work together as an international community to take collective action to close the digital data divide, while protecting our human rights?

On the practical level, cases of data for learning use are not easily classified as an opportunity or a risk, as their impacts may not fall on only one side of a clean-cut binary between positive and negative. After two years of debate and discussion, the WGD4L takes the position in this final report that data for learning is a **double-edged sword** that must be wielded with great care, purpose and intention. On the one hand, the education workforce needs accurate, timely and complete data to inform relevant and inclusive education practices and policies, but on the other, the use of education data cannot encroach on the rights of learners or deepen global learning inequalities. These key tensions are discussed in the report and summarized in the following diagram.

INTEGRATED DATA SYSTEMS

On one hand, ensuring that learning data is integrated across periods and sources can support learner mobility, smooth the transition from school to work, and improve government services responding to children vulnerability to harm, such as asylum-seeking children or undocumented youth. Data can play a crucial role in identifying schools in need of immediate supervisory support, for both operational, and teaching and learning purposes.

INTEROPERABLE DATA SYSTEMS

One of the major trends in education data infrastructures is to house information management systems on the cloud, with interoperability between platforms geared towards learning (LMS) and those with a broader scope geared towards management (EMIS). Enhancing cloud service interoperability can increase the potential for context-sensitive school management and monitoring of policy impact at the school level.

DATA-INFORMED PEDAGOGY

Teachers and school leaders may benefit from sophisticated data systems that can alert them when a learner's data displays early warning signs for course failure or drop-out. They can then combine predictive analytics with their own qualitative and contextualized insights to engage and support students or their families, reducing the likelihood of early school leaving. Learning management systems (LMS) can support teachers to identify individual and class trends in performance to differentiate instruction and prepare targeted lessons.

REGULATION

Strong legal frameworks are needed to protect learner data, regulate the involvement of commercial providers and procurement procedures, and prevent algorithmic bias, data surveillance and or misuse of education data. Strong regulation can ensure transparency, explainability and accountability in every education data use case.

MULTISTAKEHOLDER PARTNERSHIPS

Bold alliances between government and industry can accelerate the capacity development and growth of data and cybersecurity expertise in education and training systems. Partnerships between public and private sector stakeholders can lead to cost-effective, national-scale sustainable systems for the safe and effective coordination of data for learning.

DATA FOR ALL LEARNERS

The global community must accelerate efforts to bridge the digital and data divide that exists both within and across countries to ensure the benefits of data-reliant technologies and responsive data information management systems can reach all learners everywhere. Marginalized learners must be represented in datasets to enable inclusive digital tools and data-informed policies.

SURVEILLANCE OR BIAS

On the other hand, such integrated systems that provide a seamless transfer of comparable records between institutions could spur unnecessary surveillance mechanisms by tracking years of granular student data that may make negative data points (behavioural or academic) difficult to recover from, thereby influencing negative institutional and self-perception of students' success in learning systems.

COMMERCIAL CONTROL

However, this architecture also favours the market by lowering the cost of entry into existing digital infrastructures, in turn, can lead to vendor-controlled pricing of data services. It also greatly expands the amount of data that can be collected through the streamlining of data pipelines and single sign-on capabilities that can collect user data regardless of location (see, for example, Gulson *et al.*, 2022).

HYPER-DATAFICATION

Collecting data for the sake of it can create high-stress, high-stakes school environments. Limited training on data processes, combined with the pressures to continuously collect for accountability's sake, can result in teacher frustration and a decline in the attractiveness of the teaching profession (UNESCO, 2017). Technology in education should put learners and teachers at the centre to prevent it becoming a burden or an impediment to teachers' ability to teach (UNESCO, 2023*b*, p. 21).

INNOVATION

Over-complicated and opaque legal and regulatory requirements for the use of education data can prevent local companies from developing relevant digital learning solutions, preventing the growth of local education technology ecosystems that create contextualized data-fuelled digital tools for education.

CONTROL AND SUSTAINABILITY

Governments may yield their agency in steering the safe and sustainable use of education data use if essential data skills are outsourced to private providers for data harmonization, compilation, analysis of raw data, development of application programming interfaces, creation of dashboards or custom reports, and cybersecurity provisions, among others. Less than one third of countries have a sustainability clause in their procurement law.

FINANCIAL AND ENVIRONMENTAL COSTS

Data generation and analysis comes with high financial and environmental costs. Estimates indicate that affordable data is the most costly component of financing universal digital learning, with a price of $498 billion for the period 2021–2030 (UNICEF, 2021). Data centres account for a large percentage of electricity demand and produce a heavy carbon footprint.

These tensions are not easily resolved, meaning that it is of paramount importance that all education decision-makers design and review policies, practices and architectures that maximize the potential and minimize the risks of data use in education. Unleashing the potential of data use in education hangs in a delicate balance, not a binary, between benefits and risks, data-driven and data-informed, innovation and regulation, and surveillance and invisibility.

The Broadband Commission convened a working group chaired by UNESCO and composed of commissioners and experts in January 2022 to discuss the double-edged nature of data for learning along three thematic areas: (1) **infrastructure** and **architecture** of education data ecosystems, (2) data **skills** and **competence** frameworks for life and work, and (3) **governance**, regarding ethics, national sovereignty and cross-border data flows. These areas structured the work of the group and were adopted as strands of dialogue and reflection. At each monthly meeting, working group members, as well as visiting experts and institutions, were invited to present thematic cases to enrich the discussions. To build its analysis, this report explores some of these cases for how governments and industry are handling data-related issues through their domestic education policies, and bilateral or multilateral arrangements.

The culmination of this analysis is the report's **five recommendations** that the Broadband Commission WGD4L presents to policy-makers and stakeholders engaged in the education data landscape.

Overall, this report argues for the need for multisectoral, multilateral and multi-stakeholder cooperation to safely unleash the potential of data for learning to drive the improvement of education's persistent problems. Ethical and purposeful uses of data play a key role in improving school and system management, supporting inclusive and innovative learning and teaching methods, and ensuring the equitable financing of education by illuminating areas of systemic need for investment.

Five recommendations:

Develop and implement a **whole-of-government and whole-of-ecosystem vision and strategy** on the use of data for learning, grounded in a rigorous understanding of the potential opportunities, benefits, limitations and risks.

Establish a **sustainable financing strategy** for data for learning, grounded in national financial resources, that benefits learners, promotes data in education as a public good, minimizes environmental impact, and is sustained by strategic multi-stakeholder partnerships.

Strengthen critical **data literacy and skills** at all levels of the education ecosystem to facilitate improved regulation and inclusive innovation through effective implementation and monitoring of education data policies and practices.

Prioritize uses of education data that **target systemic obstacles** to universal access to equitable and inclusive quality teaching and learning, by improving the efficiency and effectiveness of education management, administration, planning and financing.

Harness **multilateralism, solidarity** and **international cooperation** to develop international standards and norms over education data regulation, literacy, cybersecurity, governance and ethics, bridge the digital divide, nurture local data capacities, and promote free open-source software to support the development and implementation of safer and more targeted tools for the education sector.

The data on education data

97 countries

lack foundational learning data to know which students can read basic text by age ten

(UIS, 2020)

US$498 billion

would be needed to make data usage affordable in learning activities for the coming decade

(Yao et al., 2021; UNESCO, 2023*b*)

In **44%** of low- and middle-income countries with pricing data, the **median cost** of 1GB of data **exceeded 2% GDP per capita**

(UNESCO, 2023*b*)

In 2022 the education sector accounted for...

5%

of all ransomware attacks

30%

of security breaches

(UNESCO, 2023b)

Globally, there is no **foundational literacy data** for

18%

of the school-age population

In Sub-Saharan Africa, this number grows to over

50%

(UIS, 2021)

60% of primary schools

50% of lower secondary schools

35% of upper secondary schools

are not connected to the Internet

(UNESCO, 2023*b*)

71% of countries in the world have **adopted laws** concerning **personal data protection**

(UNCTAD, 2021)

but only **16%** guarantee data privacy in **education** with a law

(UNESCO, 2023*b*)

53%

of countries worldwide rely on paper-based information systems

(UIS, 2020)

54%

use unique student identification numbers

(UNESCO, 2023*b*)

29%

of education ministries are unable to precisely locate schools they manage

(UIS, 2020)

89%

of 164 recommended learning platforms or products during COVID-19 were found to have tracking technologies that could or did monitor children and harvest personal data on children to send to third-party companies

(Human Rights Watch, 2022)

39 out of 42

governments that provided online education to children during the COVID-19 pandemic used digital technology in ways that violated children's rights

(Human Rights Watch, 2022)

Only 54% of countries

have mapped out digital skills standards

Many are adopting skills frameworks developed by the private sector that may prioritize technical skills over competences around data governance, ethics and social impact.

technical skills

data governance, ethics, social impact

(UNESCO, 2023*b*)

Introducing the Broadband Commission and its Working Group on Data for Learning

Since its establishment in 2010 by the International Telecommunication Union (ITU) and the United Nations Educational, Scientific and Cultural Organization (UNESCO), the Broadband Commission for Sustainable Development (Broadband Commission) has expanded both the breadth and depth of international dialogue on sustainable development, leading advisory work and advocacy for the transformational impact of broadband technologies on human lives. Working groups are at the heart of the Broadband Commission's work. With more than 30 groups to date, the Broadband Commission's working groups bring together stakeholders from all sectors to advocate for meaningful and universal connectivity, and achieve its seven Broadband Advocacy Targets (Broadband Commission, n.d.). All working groups leverage the expertise and perspectives of a unique composition of membership, comprising some of the key players in the technology industry, civil society, intergovernmental organizations, non-governmental organizations (NGOs), academia and government.

Education is a core focus of this work, and to date, the Broadband Commission has convened seven working groups on the theme of education:

1. Data for Learning (Broadband Commission, 2022*b*)
2. AI Capacity-Building (Broadband Commission, 2022*a*)
3. Digital Learning (Broadband Commission, 2021*b*)
4. School Connectivity (Broadband Commission, 2020)
5. Child Online Safety (Broadband Commission, 2019)
6. The four-year Working Group on Education (Broadband Commission, 2017)
7. Multilingualism (Broadband Commission, 2011)

Together with the 2017 Working Group on Digital Skills for Life and Work, these working groups have convened industry leaders, government officials and civil society to address prominent issues specifically dedicated to the intersection of education and technology. Building on the work and research of these groups, the Working Group on Data for Learning (WGD4L) has facilitated key dialogue on data for education and training recovery, resilience and future development, with a specific focus on data for learning.

As Chair of the WGD4L, UNESCO has organized and led monthly meetings since January 2022 to share knowledge, experiences and case studies on subjects related to education data, such as global data divides, data literacy development, data governance in the education sector, increasingly integrated and interoperable data ecosystems, and the ethics of using data-fuelled technologies in education and training.

The WGD4L, through its collaborative work and this report, aims to strengthen the foundation for ongoing discussions of the dynamic education data landscape by filling existing information gaps and mapping the different issues at stake in this debate, including (1) examining the **potential assets** and **potential dangers** of data for learning, (2) understanding the **drivers** and **barriers** to investment in data for learning, and (3) **anticipating the development** of the data ecosystem and its impact on education.

Scope of the report

The WGD4L set a broad scope, situating data for learning within broader **lifelong learning** and **right-to-education** perspectives. As such, learning should be understood not solely as a formal activity affecting children within the walls of a school building. Likewise, learners are not only children, but rather people of all ages and localities engaged in educational and training opportunities.

The seventh International Conference on Adult Education made clear that the global community is taking steps towards affirming the right to education throughout life (UNESCO Institute for Lifelong Learning, 2022). Data play a critical role in supporting individual learning pathways for adult and tertiary-level learners, especially with regard to supporting reskilling and upskilling for the future of work. However, given that school-based learning is the primary pursuit of most children and youth, the WGD4L gives specific attention to the rights-based issues affecting young people's relationship to learning data.

The collection and use of data from learning spaces need to be aligned with strong national data policies and regulatory frameworks that ensure data privacy in education by law. However, this rights-based perspective on education data extends beyond concerns regarding the privacy and security of personal data. It also implicates the need to mitigate risks relating to data profiling, deterministic algorithmic interference, environmental impact, and commercial interests that may divert control over the education data ecosystem away from public authorities and undermine the affordability of data services and their applications in learning spaces.

While data-fuelled technologies offer possibilities for empowering educators, learners, schools and education policy-makers, global gaps in access to electricity, broadband networks and digital technologies in schools are significant contributors to persistent and widening disparities between educational outcomes in low- versus high-income contexts. While governments in well-connected contexts may be able to use data to improve the efficiency and effectiveness of system management to deliver on all learners' right to quality and inclusive education, disconnected contexts are flying blind without access to information on student behaviour and outcomes.

Objectives and structure

The main objective of this report is to provide an overview of the key debates surrounding data use in education, and stress the need for multi-stakeholder efforts to ensure timely, transparent and trustworthy collection of data to be used for a specific purpose, such as to improve learning, teaching, administration or strategic management. To accomplish this objective, the report provides recommendations to policy-makers to establish strong, cross-cutting education data policies and governance architectures that are responsive, resilient and ready for an increasingly digitalized future.

To explore the double-edged nature of education data, the WGD4L has focused on **three strands** related to data for learning: (1) data infrastructure and learning ecosystems; (2) data skills and competence framework for life and work; and (3) ethics, governance, national sovereignty and cross-border data flow regulation.

Accordingly, the report is organized into five parts:

→ **Part 1** explores, defines and refines data, in particular data for learning, discusses why data are important for learning, and introduces the idea of data for learning as a double-edged sword.

→ **Part 2** then explores the education data ecosystem, looking at the modernization of data infrastructures, the link between data for administration and data for learning, and uses and users of education data across integrated and interoperable education data systems.

→ **Part 3** discusses data literacy and defines appropriate competencies across the education data ecosystem.

→ **Part 4** investigates important considerations for education data governance, including financial and environmental sustainability, ownership, and ethics.

→ **Part 5** synthesizes the WGD4L discussions of the potential benefits and risks of data for learning, culminating with recommendations designed to guide the development of policies for data use in education.

Finally, the report concludes by anticipating challenges to the future of data for learning, and exploring areas for further analysis, data collection and policy dialogue.

1

Defining data for learning

Data ecosystems are expanding and evolving, propelled by the digital transformation of the Fourth Industrial Revolution and the acceleration of digitalization due to the COVID-19 pandemic. Information industries, powered by the technological capabilities of cloud computing, artificial intelligence (AI) or the Internet of Things (IoT), increasingly influence decision-making across contexts. The data revolution has rapidly changed how and what services are produced and delivered, not only within industrial sectors, but also within social sectors such as health, social security and education.

The share of the population using the Internet has grown exponentially over the past decades, with 16 per cent of the population using the Internet in 2005 growing to 66 per cent by 2022 (ITU, 2022). However, it is the arrival of cloud-based data management services, and data-reliant technologies such as AI and block chain, in parallel with advances in global telecommunications networks, that have significantly amplified the volume and complexity of data flows.

> 90 per cent of the world's data was generated in the last two years alone. In 2010, 2 zettabytes were generated annually. Estimates predict this will rise to 175 zettabytes of data annually by 2025. This represents an almost 88-fold increase in just 15 years.

There has been a huge surge in the production of digital data globally, including cross-national data flows, over the past 15 years, driven by widespread personal use of the Internet and email, advances in data-intensive technologies, and the interconnectedness of the global economy. Data from April 2022 shows that **almost 250 million** emails are now sent every minute. Across 24 hours, that adds up to **333.22 billion** emails (Duarte, 2023). The amount of global data generated annually has grown from 2 zettabytes in 2010 to a predicted **175 zettabytes** in 2025. It is also estimated that 90 per cent of the world's data was generated in the last two years alone (Statista, 2022).

In terms of traffic, although a unified measurement of cross-border data flows does not yet exist, Internet traffic can be used as a useful proxy. Looking at the numbers, it then becomes clear that by the end of the 2010s, an already-existing upward tendency in Internet traffic was accelerated by the COVID-19 pandemic, when many activities moved online (United Nations Conference on Trade and Development [UNCTAD], 2023). However, this increase in Internet traffic is not equally distributed across and between countries,[1] with traffic predominantly concentrated on two main east-west routes, between North America and Asia and North America and Europe.

Although global data generation is a rising tide, **2.7 billion people are still offline today**, and 90 per cent of these people live in developing countries (ITU, 2022). Under the digital divide lies a deepening data divide, which is defined as the gap that exists between those who can take advantage of the opportunities offered by digital data, and those who are left further behind. In the education sector, connectivity plays a key role in this global gap, with **one in four primary schools** globally having no electricity, and **60 per cent** of primary schools, **50 per cent** of lower secondary schools and **35 per cent** of upper secondary schools **not** connected to the Internet (UNESCO, 2023*b*). Beyond electricity and Internet access, a lack of sufficient access to devices and limited digital skills are additional barriers to purposeful data use in education.

Moreover, there are significant differences in Internet affordability worldwide, often with the poorest people paradoxically paying the most expensive mobile data fees. Globally, there is a **30,000 per cent** difference between the cheapest data price and the most expensive (Ang, 2020), with India ranking the cheapest at US$0.09 per gigabyte, and Malawi ranking the most expensive at US$27.41 per gigabyte.

Therefore, this report is grounded in an awareness of the persistent **digital divide** and **data divide** between and within countries, which create global imbalances in the potential to leverage data to improve equitable and inclusive learning. Disconnected learners, marginalized learners, and women and girls risk being under-represented in datasets. The

1 This statement refers to openly available information from TeleGeography, the largest provider of data and analysis on long-haul networks and the undersea cable market. More information is available for subscription; it is therefore possible that more detailed statistics exist, but are proprietary.

misrepresentation or invisibility of certain learners in these datasets may result not only in ineffective data-reliant educational technology (EdTech) tools, but also in a reproduction of broader social inequalities. Furthermore, contexts with wider digital divides may also have limited legislative protection of learner and learning data, as well as lower levels of data literacy, which could endanger the data sovereignty and security of both countries and individuals, in particular the most vulnerable.

The international community is not blind to the inequality in the digital data revolution. Secretary-General of the United Nations António Guterres raised concerns in the 2021 **Our Common Agenda** report (United Nations, 2021) that issues of data are deeply intertwined with the digital divide. The report calls for international cooperation to connect all people and schools to the Internet, avoid Internet fragmentation, protect data, apply human rights online, introduce accountability criteria for discrimination and misleading content, ensure the digital commons is a global public good, and promote the regulation of AI. This call will be formally agreed upon in the form of a **Global Digital Compact** in September 2024, which is expected to "outline shared principles for an open, free and secure digital future for all" (United Nations Office of the Secretary-General's Envoy on Technology, 2023).

To advance along the Secretary General's roadmap for digital cooperation, the UN-endorsed multistakeholder **Digital Public Goods Alliance** (DPGA) and UNESCO and UNICEF's **Gateways to Public Digital Learning** initiative are actively championing digital public goods and digital public infrastructure around the world.

> **United Nations Member States adopted the Declaration on the Commemoration of the Seventy-Fifth Anniversary of the United Nations (A/RES/75/1), which contains a pledge to improve digital cooperation through a Global Digital Compact, to be agreed upon by the United Nations, Governments, the private sector and civil society at the Summit of the Future in September 2024.**

Despite this progress in international digital cooperation, not all countries have adopted the frameworks and legislation needed to address concerns connected to data privacy protection, regulatory control, data security, affordability, and integrity. In fact, 15 per cent of countries have no legislation in place at all, and 5 per cent have no data on the subject (UNCTAD, 2021). Most of these countries are in Africa and Asia, with less than half of the least developed countries having adopted such legislation. Even for the 71 per cent of countries that do have privacy legislation, regulations on data are not always specific to educational contexts, nor to children. In fact, only **16 per cent** of countries guarantee data privacy in education in law (UNESCO, 2023*b*).

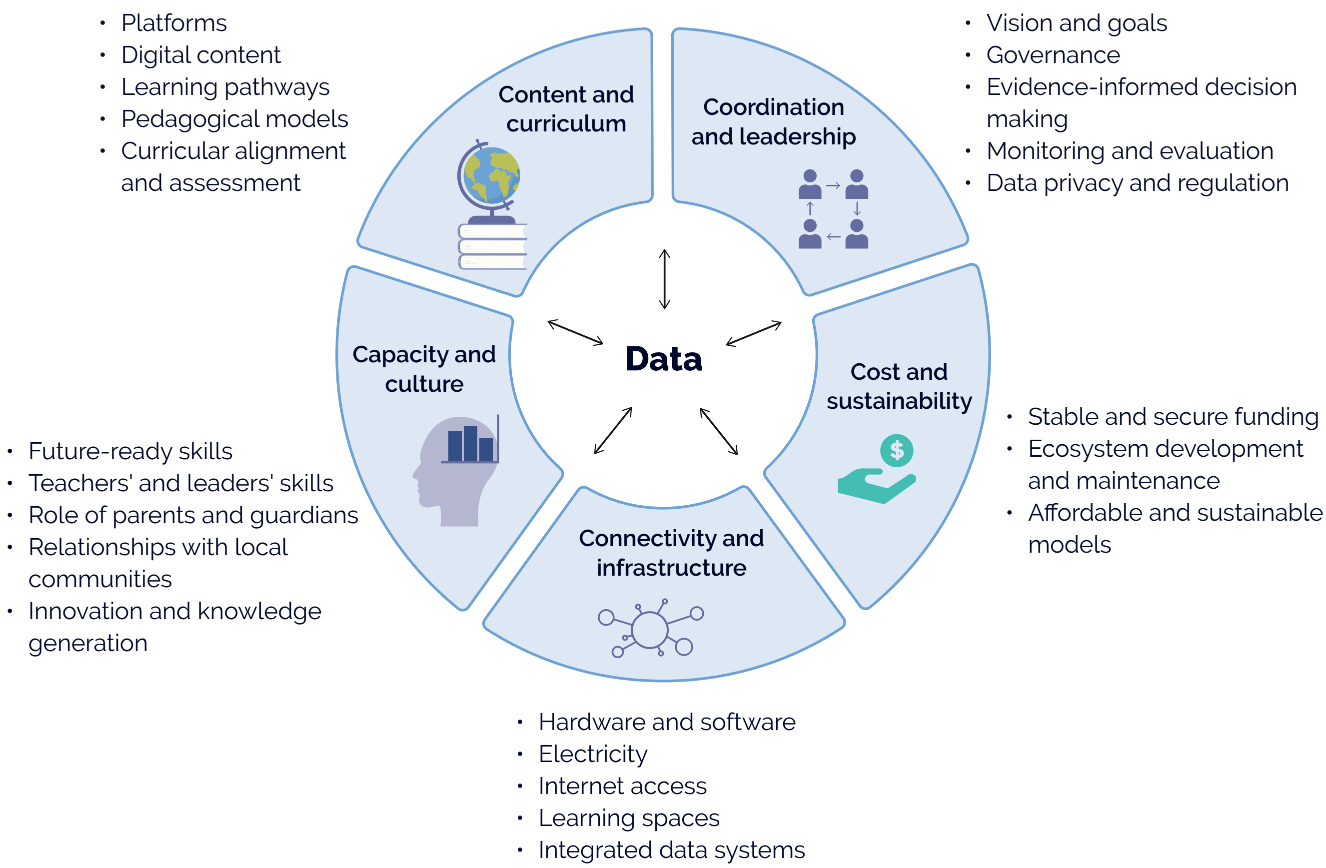
> **Today, over 71 per cent of countries in the world have adopted laws concerning personal data protection, with 9 per cent in the process of development and 15 per cent with no legislation in place; 5 per cent of countries have no data on such legislation. Most countries without such legislation are developing nations.**
> (UNCTAD, 2021)
>
> **For the education sector's specific needs, only 16 per cent of countries guarantee data privacy in education with a law.**
> (UNESCO, 2023*b*)

To move international work forward on how these normative instruments apply to education specifically, UNESCO launched the **Evolving Dimensions of the Right to Education initiative**, which examines new challenges and trends facing the right to education, including in the digital sphere. Protecting learner data and expanding equitable access to digital services will require sustained action to share knowledge and establish norms and standards for data security and use (UNESCO, 2022*b*). This will only be possible through public-private partnerships (PPPs) between industry and government, for which the Broadband Commission has long served as an exemplar.

In the same spirit of multi-stakeholder solidarity, UNESCO convened the **Digital Transformation Collaborative** (DTC) during the Transforming

Education Summit of 2022, a group of partners representing EdTech, international organizations, telecommunications, NGOs, funds and foundations committed to collaboratively steering digital transformation in education towards equity, quality and sustainability. The DTC created a common framework for system-wide digital transformation in education called the **Five Cs Framework**. This unpacks the core components of digital transformation in education needed to deliver a positive impact on learners, teachers and all education stakeholders, and is visualized in Figure 1.

Figure 1. Five Cs Framework for Digital Transformation in Education



Source: UNESCO (2023*d*).

Taken together, the framework above could be seen as a wheel to propel progress towards equitable, inclusive technology-enabled education. For sustainable, meaningful benefits to be experienced system-wide from using digital tools in learning settings, the following five areas must be considered:

- **Coordination and leadership** refer to having a clear vision, goals, and evidence-based plans for implementation of plans to use digital technology in education, which requires strong cross-sectoral relationships, multi-stakeholder partnerships, legal and regulatory frameworks, and strong monitoring and evaluation strategies to realize the vision and deliver on the plan.

- **Content and curriculum** are the platforms, digital content, learning pathways, pedagogical models, and curricular alignment and assessment supported by an education system.

- **Capacity and culture** indicate the future-ready skills of all education stakeholders, including teachers, leaders, parents and guardians, leveraging relationships with local communities.

- **Connectivity and infrastructure** are the hardware, software, electricity, Internet access, learning spaces and data systems of the education system.

- **Cost and sustainability** concern the funding, development, maintenance and cost-effectiveness of the financing models for tech-enabled learning in the education ecosystem.

Throughout these five components are **data**, as depicted in the centre of the wheel with bidirectional arrows. This placement implies that not only does a system need education data to understand where there may be bottlenecks or gaps within these five areas of digital transformation, but also that new data emerges as a result of the process of digital transformation in education. As parts of the learning, teaching, management and administration processes are digitalized, new digital data is produced, which can then be used to improve management, administration, financing, teaching and learning. In essence, the model illustrates that data is a crucial lever for and product of digitalization in education.

While there is a sense of urgency in advancing education systems across all components of digital transformation in education, the global landscape for the governance of data is still fragmented, especially within the education sector, with countries and regions adopting different policy approaches to regulate and safeguard data flows across their borders. Without smooth and secure flows of education data to identify gaps in these five areas of digital transformation, to know which children are learning and what they are learning, and to make data-*informed* decisions to improve system management, the potential of digital technology to target education's systemic challenges cannot be leveraged.

As such, the digital and data divides in education are intrinsically linked, but concerted efforts to close the data divide – regarding both access to timely, transparent and trustworthy data, and the skills to understand and use that data safely and purposefully in education – are less evident or discussed than those regarding digital technology more broadly.

Countries struggle to find common taxonomies to discuss data in education as the basis for collective agreement on data protection and data flows. The various taxonomies used to classify data categories are sometimes based on different criteria. For example, data may be collected for commercial or governmental purposes, used by the private or public sector, and may be instant or historic, sensitive or non-sensitive, or personal or impersonal (UNCTAD, 2021). Agreeing on a common language with which to discuss the double-edged nature of data for learning has not been prioritized within the education sector.

> **Different understandings of key data terms and approaches can hold back multilateralism and country agreements on what types of data – including education data – might be tracked across borders for different purposes supporting the common good.**

As a starting point, this report begins by outlining an operational definition of data for learning, accompanied by a taxonomy specific to the education sector, to guide its analysis of the benefits and risks of data use in education.

An introduction to digital data

Before examining the potential assets and dangers of data for learning, it is important to begin by defining what we mean by data. The concept of data is often used synonymously with information or evidence, but given the volume, variety and velocity of digital data sources available in today's world, the word deserves a precise definition. In this report, we use data to mean sets of discrete items of information such as numbers, text, images or sounds that are collected, cleaned, formatted, stored and shared, and used for analysis, calculation, inference and application.

Cross-national data flows have long played a key role in diversifying and accelerating progress across economic, financial, health and cultural sectors. The smooth operations of global trade and commerce, and the expansion of the global financial system, were built on the cross-border exchange of economic and financial data shared between businesses, governments and logistics providers in different countries. Like the game-changing use of steam power to turn the wheels of the Industrial Revolution, new uses of data have catalysed a data revolution driven by a data economy where value is derived from accessing, gathering, organizing and controlling information.

The digital economy and e-commerce have exploded on the back of global digitization trends and technological advances, with the transactions of cross-border e-commerce estimated to exceed US$2 trillion in 2023 (Digital Nation Staff, 2022). This revolution is fuelled by recent changes in scale, power and skills. The vast amount of data that is constantly being created by emergent web, mobile and digital technologies in virtually all areas of human activity, the powerful computers and data analytic techniques now available, and the growing numbers of individuals with basic or advanced data skills, have sharply reduced the costs of collecting, storing and using data. These vast amounts of data being created and collected are often referred to as "big data", the attributes of which are described in the Table 1.

Table 1. The "V attributes" of big data

Volume	The size of the dataset is very large.
Variety	The different types of data are generated from multiple sources, needing to be cross-referenced and combined in order to be fully exploited.
Velocity	The data may be generated at a rapid rate.
Veracity	The data may be incomplete, influencing the precision of inferences made from it.
Volatility	The data being collected or inferred may become less relevant over time.
Value	The ability to extract value from such data while complying with given time, human and technical resource constraints.

Source: Adapted from du Boulay *et al*. (2018, p. 269).

With this big data revolution has risen a recognition that new approaches to data production and use can yield more dynamic insights that were previously unattainable with slower, traditional methods of information management. Big data also "refers to things that one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more" (Mayer-Schonberger and Cukier, 2013, p. 6).

However, "big" is often conflated with "beautiful" or "better", usually without explicit justification:

> *We tend to prioritize large data sets instead of small data sets. We believe bigger data sets will tell us more and will provide more accurate information than smaller data sets. But that is not always true... it's not binary – you can have big and small. Both tell you something, but neither tells you everything. The second thing is we tend to think that collecting more data will solve the problem. Here's a problem, let's collect some data on it and then we'll get closer to a solution. But in some cases, that is actually not the best thing to do … We call this the paradox of exposure*. (FEED Staff, 2020)

Big data is increasingly described as the fuel that drives businesses and organizations forward. Countries around the world are investing heavily in their data-processing infrastructures in the hopes that they can transform the ever-growing piles of big data into actionable, real-time information. However, without being analysed and transformed into usable information, big data are crude and have little value. Moreover, using this processing power to transform data into useful information comes with rising costs – both financial and environmental. Indeed, the value and benefits of small and localized datasets should not be ignored in the face of big data.

The lifecycle of data – creation, cleaning, processing, analysing, transferring, storing, archiving or destroying – is determined by humans and comes with a substantial carbon footprint and electricity bill. Thus, the particular social, economic, technological, political or environmental perspectives baked into big datasets and the applications trained atop them should be accounted for when conclusions are drawn from them.

In the education sector's efforts to ensure agency, awareness of data use, and bias-free datasets and algorithms, it is necessary to briefly outline a typology of data used in education: **metadata**, **digital transactional data**, **aggregate data** and **synthetic data**. These four types are explained in Table 2 and Figure 2, along with their potential uses in educational contexts.

Table 2. Types of digital data and common uses in learning spaces

Type of data	Description	Examples	Potential uses for learning
Metadata	Data about data. Metadata often is described as information about data, encompassing details like collection methodology, context, variables and relationships, crucial for understanding and managing datasets.	**Traffic** data, such as digital platform logs that track the date, time and recipients of digital communications; **identity** data, such as IP addresses, device numbers or smartphone sensor data; **location** data such as the reception area of SMS or telephone contacts.	Often used to operationalize and understand knowledge, cognitive strategies and behavioural processes, in order to personalize and enhance instruction and learning. Traffic and location metadata can reveal private or sensitive information on the persons concerned, and is therefore of particular importance in discussions of protecting learner rights in digital environments.

Transactional	Information captured and recorded about an event, which typically includes time, numerical values and references to one or more objects.	Any data expressions directly generated by users, e.g. journalling, social media posts and online discussion forum comments.	Often used by institutions to understand how users interact with a website (time stamps on site traffic, popularity of topics, language used in comments, etc.). Often analysed with natural language processing techniques to relate linguistic features to cognitive, social, behavioural and affective processes.
Aggregate	Individual-level data from multiple sources that disappears from individual isolation to be combined and summarized for the purposes of examining trends, making comparisons, reporting, etc.	Institutional data, student demographic data, graduation and enrolment rates, school standardized test performance scores, etc.	Often used to inform administrative decision-making; can be used to improve course enrolments or student engagement through data analytics, as well as cases of AI-powered use like course guidance systems and predictive systems.
Synthetic	*Can exist at any of the above levels.* Data that mimics real-world data, generated using sophisticated AI models to create whole new datasets from scratch.	Any dataset that does not exist in the real world, which can be applied to mimic any type of data, from insurance data (Hann, 2021), to self-driving vehicles (Behzadi, 2021), or even patient health care records (Walonoski *et al.*, 2017). Developers can train cars on virtual streets and can supply synthetic human faces on demand.	Often used to supplement or supplant real-world data with "better", "cheaper" or "bigger" datasets constructed using AI (Koperniak, 2017; Lohr 2018) to (1) lower the cost of developing helpful AI algorithms; (2) improve the diversity of datasets to counter implicit bias or invisible data in "real" data; and (3) provide better privacy protections and lower the use of sensitive personal data, such as children's data.

Source: Authors.

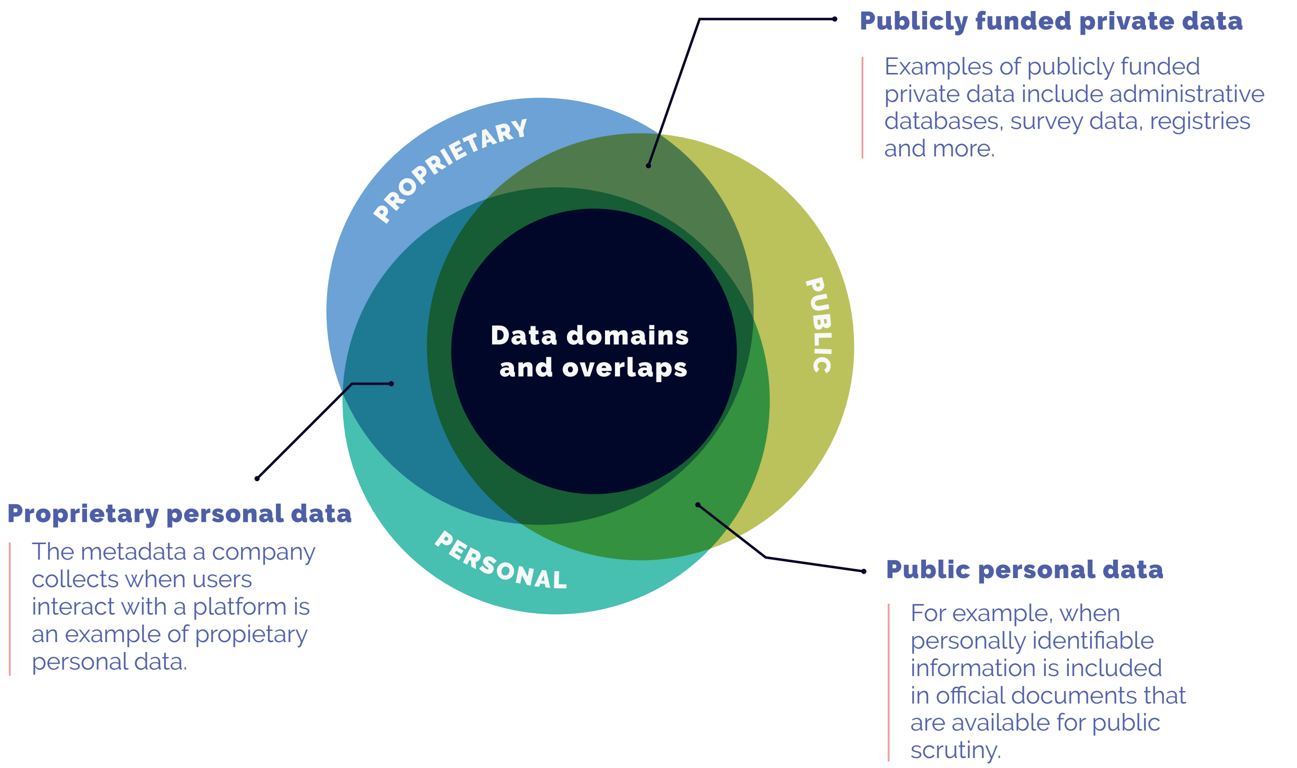
Data collection, exchange and analysis all often involve better understanding of the various domains, as well as the way the data were actually generated. For data governance frameworks to be applicable across various learning contexts, it is necessary to distinguish between the three domains of data:

- The **personal** domain covers all data relating to an identified, natural or identifiable individual (personal data) for which data subjects have data rights.
- The **proprietary** domain is typically protected by intellectual property rights (IPR) (including copyright and trade secrets), or by other access and control rights (provided by legal contracts, cybercriminal law, etc.). There is typically an economic interest to exclude others.
- The **public** domain covers all data that are not protected by IPR or any other rights with similar effects, and therefore lie in the "public domain" (understood more broadly than to be free from copyright protection); certain types of such data are thus free to access and reuse.

These domains often overlap in real-world scenarios and are typically subject to different data governance frameworks. For instance, privacy regulatory frameworks typically govern the personal domain, while the proprietary domain is mostly governed through contractual frameworks, or in some specific instances, by IPR. In the case of data co-creation, multiple stakeholders are often involved in the contribution, collection and control of data.

Given these overlapping domains and multi-stakeholder involvement, as illustrated in Figure 2, the governance and regulation of data is considered legally complex. This is especially true where cross-border data flows are concerned. Currently, data-related rights vary significantly between countries, which adopt different approaches to personal and proprietary data for individuals and companies. Understanding the interplay between personal, proprietary and public data can help determine the level of awareness users should have of data processes, their rights in accessing and using data, and the jurisdiction of data generation and collection to identify the applicable legal and regulatory frameworks (Organisation for Economic Co-operation and Development [OECD], 2020).

Figure 2. Data domains and overlaps



Source: Authors.

Why is data important for learning?

The world's eyes have recently been opened to the vast data gaps that render invisible the needs of many vulnerable learners across regions. These gaps exist in contexts across the digital transformation spectrum. In high-income countries, only a narrow range of learner data may be collected, while in low- and middle-income countries, learning data may be collected infrequently or not at all (UNESCO Institute for Statistics [UIS], 2021).

In July 2021, the United Nations Children's Fund (UNICEF), UIS and the World Bank joined forces to make a **Learning Data Compact** to ensure that all countries have at least one quality measure of foundational learning by 2025. This is one of many recent examples of international solidarity that demonstrate the global community's growing recognition of the importance of data in improving the quality of learning, teaching, administration, management and governance.

> **As of 2021, 97 countries are missing foundational learning or schooling data. Globally, there is no foundational literacy data for 18 per cent of the school-age population – in Sub-Saharan Africa, this number grows to over half of all school-age children.** (UIS, 2021)

Data plays a key role in determining progress towards the **Sustainable Development Goals (SDGs)**, in particular in measuring SDG 4 on quality education, to ensure inclusive and equitable quality education, and promote lifelong learning opportunities for all. In countries with higher capacities for data production and use, data can smooth the divisions and transitions between central, subnational and school levels. When used ethically and effectively, data can help administrators, teachers and learners alike make evidence-based decisions, plan strategically towards long-term goals, adjust strategies as situations evolve, encourage flexible pathways, and support context-relevant innovation in education. Moreover, data-based technologies can be particularly useful for scientific research, constituting datasets that can drive innovation and knowledge generation.

Distinguishing learner data from learning data

Learner data describes any data that is personally identifiable and can be gathered about learners by teachers or products in the classroom (both qualitative and quantitative) through observations, assessments and testing, and, increasingly, through EdTech applications. Learner data can have benefits for personalized instructional practice, helping to fine-tune classroom pedagogies and generating a better level of individualized differentiation for learners. Learning data, however, does not necessarily have to be personally identifiable, and focuses on learning processes and outcomes that can inform next steps in learning journeys and appropriate resources or support. This is a form of transactional data (as described in Table 2). Combining learning data with data on other dimensions of education – such as the local context, teaching and learning environments, and learner characteristics – can shed light on factors that most affect learning outcomes. By revealing gaps in student achievement and service provision, learning data can be used to better support learners, specifically those groups which are currently underserved or are underperforming, and to hold education systems accountable for the use of resources.

Table 3. Sources and examples of learner and learning data

Source	Example
National sources[2]	• Large-scale learning assessments • Education management information systems (EMIS) databases • Multi-year sector plan documents, medium-term expenditure frameworks, and operational documents (operational plan documents, budgets, mission reports and minutes of coordination meetings) • Financial data collected through financial management and reporting systems
National school data (e.g. from the central pupils' record)	• Name • Date of birth • Gender • Ethnicity • First language • Special educational needs and disability • Home address • Unique student number
Partner-facilitated sources	• Surveys and rapid assessments in the education and other sectors (including health, water, sanitation and hygiene, etc.) • Outcomes of multi-stakeholder joint monitoring and review exercises • Outcomes of decision-making within a multi-stakeholder policy dialogue forum • Data available through partner reporting • Data available through education and learning ecosystems
Data generated for safeguarding and special service provision	• Health data (from health and safety management) • Demographic data (for looked-after or vulnerable children) • Online activity monitoring
Data generated by learning tech for management	• Lesson and homework delivery • Sometimes biometrics data for accessing facilities such as libraries or cashpoints
Data collected directly from teachers and learners	• Qualitative data routinely collected by teachers through their teaching practice and multimodal assessment of learner progress • Data collected directly from learners through use of EdTech in the classroom • Data collected from stakeholders and beneficiaries through WhatsApp networks, social media groups, online and offline communication networks groups, etc.

2 In many countries, local education agencies and bureaux are also active sources providing accountability, enrolment and other demographic data.

Elements of learning that data systems may not be able to capture	• Elements of learning that data systems may not be able to capture reliably and ethically • Understanding, curiosity, imagination, creativity, thoughtfulness and collaborative processes of learning • Student frustrations, disappointments, missed learning opportunities, anxieties about learning, chilling effects of surveillance, and workarounds to proctoring services • Learner social-emotional interactions, isolation and engagement/ disengagement • Data on informal, self-directed learning • Peer interactions, both within and outside classrooms • Social processes involved in navigating the learning spaces and material
Data generated by learning technologies for learning and assessment	• Lesson and homework delivery • Online learning, including attendance and absence and resulting metadata (e.g. IP address, device information) • Assessment and testing results • Behaviour traits data, for measuring engagement and usage • Data collected through learning management systems (LMS) • Data collected through online portals and open educational resources

Source: Authors, adapted in part from Livingstone and Pothong (eds.) (2022) and Global Partnership for Education (2021).

A decade ago, data were often described as one of the most prized commodities, triggering a trend in collecting copious amounts of data, sometimes without a distinct purpose. Today, much of the world's data focus has shifted from data collection for the simple sake of amassing more data, to data collection for specific use within a programme or existing system. This trend is known as **data minimalism**, which refers to the idea of collecting only what is necessary for a specific purpose, such as the training of a context-specific digital learning programme. For example, **Kennisnet**, one of the public information technology (IT) organizations for education in the **Netherlands**, estimates that big data is approaching a plateau of productivity that would render it obsolete, as it is only valuable when applications can derive meaning from it (Van Wetering *et al.,* 2020, p. 40).

Collecting data just for the sake of it may be meaningless, and perhaps illegal, as regulatory frameworks have strengthened over the past few years to limit – or even forbid – access to personal learner data, and to require clarifications of intended use. While this may protect learner data, it may also restrict anonymized learning data from being shared as a common good to train other systems or inform changes outside of the context in which it was collected. In institutions, this could prevent learning spaces from centralizing their learning data (and not personal *learner* data) to improve wider data system interoperability, or to train locally relevant digital solutions that address distinct issues.

What learning data can we *not* collect?

Digital data collected in education are **partial** and cannot present a complete picture of all teaching and learning processes. Increasingly, they present interaction data from electronic systems but not other aspects of learning, such as reading, creating projects, talking to teachers, etc. However, the partial data that is collected and analysed can come to be taken as representing the whole, such that if a phenomenon does not appear in the data, it is either unimportant or effectively does not exist. EdTech is only capable of capturing data when the learner interacts with the EdTech system, which often constitutes only a **small part of the learning experience**.

For example, no data are captured when the learner is involved in collaborative learning or project-based learning; reading a paper book or involved in learning outdoors; writing a poem, painting a picture or performing in a play; learning work-oriented skills such as mechanics, hairdressing or hospitality; or engaging with one another, for example in discussions between themselves, or with their teachers. All these factors, although complex, are tacitly considered by most experienced teachers during their day-to-day interactions with their learners, but remain as **qualitative and contextualized data** that may be hidden or ignored, and thus unavailable to analyse.

Box 1. The importance of qualitative data: A case from India

Data existing in a qualitative/disaggregated form, especially involving classroom interactions, the level of student involvement, their interest and anxieties, etc., can be effectively used by teachers and schools to improve the learning process. This type of data can also positively influence the design of assistive technologies aimed at supporting the needs of all learners, especially marginalized learners and learners with disabilities. Evidence of this can be found in the Ludic Design for Accessibility framework, designed at Microsoft Research India, and currently being adopted by the Computational Thinking for Persons with Visual Impairment (VICT) Project. The VICT Project aims to make science, technology, engineering and mathematics education accessible to students with visual impairments in India (Ludic Design for Accessibility, n.d.; Microsoft, 2019).

The play-based pedagogical approach, implemented jointly by the non-profit enterprise Vision Empower, Microsoft and the Centre for Accessibility in the Global South at the International Institute of Information Technology Bangalore, has introduced a range of accessible games. These include traditional games that use tangible artefacts and specially designed accessible card games to help children with blindness develop foundational numeracy and computational skills (Vision Empower Trust, n.d.).

Throughout the academic year, detailed observations and rich qualitative data are gathered from classrooms, with analysis of this data at local levels informing the introduction of novel methods into the learning process, placing a greater emphasis on children's participation and enjoyment. Including this qualitative data in the analysis of the programme has helped to drive success and scale the initiative from partnering with three schools in the 2019/20 academic year to over 100 schools at the time of writing.

> While data-*driven* policies tend to be based directly on a surface reading of the data, data-*informed* policies are inferred from a contextualized and critical interpretation, one that balances the data with human behaviour and shared principles.

If subjects that are more visible in data systems are those which are more easily quantifiable, there is a risk that over time, certain subjects – for example, interpretive subjects such as the arts and the humanities – may receive less attention or value within education systems. Moreover, data captured about the transition from education to the world of work, and **work-based, informal and non-formal lifelong learning**, are not often integrated into information management practices in education systems. Fostering a deeper understanding of what learning experiences best lead to individual and societal flourishing is impeded if data from outside the formal education sector are invisible within education information systems.

A double-edged sword: Balancing benefits and risks

Unleashing the potential of data use in education hangs in a delicate balance between benefits and risks, innovation and regulation, surveillance and invisibility. These tensions are not easily resolved, preventing a clean-cut binary between positive and negative use cases.

At the conceptual level, increasing classroom data collection in education contributes to a parallel increase in the quantification and analysis of classroom activities. Academic scholars call this the **"datafication"** of education, which indicates a wider methodological shift in education policy-making that seeks to simplify the complex socio-technical interactions within an education system to dashboards and computational models drawn from large, quant-heavy datasets (see, for example, Williamson, 2016; Livingstone and Pothong (eds.), 2022). EdTech could contribute to a reductive view of learning that values only that which can be numerically measured, tracked and standardized, undermining the belief that all learning, including digital learning, is socially situated. Selwyn (2015, p. 72) calls this the "recursive" effect, whereby "data analysis begins to produce educational settings, as much as educational settings produce data." There are well-documented examples of this effect in action, including schools that have become sights of surveillance of both teachers and learners in the name of transparency and accountability.

At the practical level, every successful example of data for learning must also contend with wider concerns posed by the growing trend towards datafication and digitalization in education (for specific discussions on these key concerns, see Part 4). Given the complex interconnectedness of education system functioning, each decision related to data for learning practices can have unintended ripple effects on other parts of the system. As shown in the case of Australia studied by Carroll *et al.* (2022), hyper-datafication can create high-stress and high-stakes school environments where teaching and learning are positioned as competitive pursuits, as opposed to gradual and lifelong learning processes of development.

Overall, the short- and long-term impacts of datafication in education cannot easily be classified as either an opportunity or a challenge, as they may be neither entirely positive nor negative. This report takes the position that data for learning lives within this binary of the **double-edged sword**, and must be wielded with great care, purpose and intention. Given this plural impact, it is of paramount importance that all education decision-makers design data policies, practices and architectures that maximize the potential and minimize the risks of data use in education.

On the one hand, purposeful data use in education could contribute to the detection and monitoring of various inequalities in education, which could then lead to improved resource allocation and, ultimately, more inclusive education systems. On the other hand, the long-term effects of datafication could shift pedagogical decision-making away from classroom teachers and towards those with the power of big data analysis. Pressures on data-*driven* rather than data-*informed* instruction could lead to the undervaluing of essential social interactions in the classroom that ensure learners and teachers remain humanized rather than turned into data points devoid of context.

Real-time data systems can bring visibility to learning trends or learner traits that teachers may not have seen, that then improve their differentiated instruction. However, overreliance on such systems could also increase the reliance on third-party data service providers, thereby augmenting commercial influence over education provision and decision-making and introducing new threats to data security and privacy. With the necessary guard rails, private sector involvement in the provision of EdTech could complement the value of education as a public good and a universal human right. As some of the primary stakeholders in this EdTech expansion are philanthropic foundations, education publishers, venture capitalists and technology companies (Regan and Khwaja, 2019), it is important that the public steers the long-term directions of data use in education.

Data-*driven* technologies are not simply unbiased teaching tools, but part of wider social systems within countries at vastly different stages of digital data culture and capacity development. The uses of advanced data analytics in education remain a double-edged sword. As such, to successfully develop inclusive and equitable education ecosystems, all education decision-makers must be cognizant of the inherent partiality of data, which is too often assumed to holistically represent individual identities. Instead, this data needs to be seen for what it is: part of a complex, socially situated whole.

Today, many digital tools designed to capture and analyse myriad dimensions of digital education data are increasingly integrated into larger data systems that seek to piece together many parts to approach an understanding of the whole. The next chapter will explore the increasingly integrated and granular nature of education data systems, such as EMIS, that mesh data from multiple sources through complex and comprehensive data architectures.

Box 2. Beyond the binary: Data in generative AI and its applications in education

The education sector has recently been rocked by the rise of generative AI tools. Although the conversations about generative AI and education data are distinct, they share many of the same critical considerations, as generative AI requires a system comprising massive datasets and set of algorithms to operate. In the case of sophisticated chatbots like ChatGPT, this system is a large quantity of text data. Using mathematical probabilities, a large language model (LLM) is then trained on this large dataset to read, recognize, summarize, translate, predict and ultimately generate new text. For example, the LLM of ChatGPT is called GPT-3, which is one of many LLMs.

Like any tech-enabled tool, generative AI applications can have both positive and negative effects on education. Discussions around these impacts have been prevalent and polarizing since the release of ChatGPT in November 2022. Some commentators speak of the promise of personalized learning pathways, streamlining administrative tasks, improving data management and education decision-making, or bolstering school cybersecurity and learner data protection in an increasingly digitalized education sector. Others fear these tools will replace teachers, encourage plagiarism, and amplify existing biases and discrimination in education systems through biased training datasets.

Similar to the double-edged sword of data use in education, education stakeholders have the responsibility to move beyond the binary of opponent or proponent of generative AI. Rather, they should move towards aligned approaches regarding the practical application of generative AI tools in teaching, learning, administration and management, and how to collectively create the necessary ethical, regulatory and human-centred educational environments that would enable positive use cases of generative AI while minimizing its risk factors. For more on the future implications on the wide availability of such tools in the education sector, see Part 5.

2

Examining education data architecture and infrastructure

Given that it is internationally recognized that the right to education starts from birth and continues throughout life, many education systems are moving beyond rigid stages of schooling and towards supporting learning pathways on a lifelong continuum. The lifelong learning agenda is increasingly integrated into laws, policies, budgets and education sector governance architectures. Beyond formal education, data are increasingly collected on the number of learners in vocational or higher education institutions, as well as about young people who are not in employment, education or training, or those who participate in non-formal or lifelong learning opportunities. Moreover, integrating digital technology into learning spaces increases the potential for collecting large amounts of granular learning data about individuals when they are engaging with an electronic data-gathering system. In response to these changes, information systems are in a dramatic state of evolution to meet the growing demand for sophisticated and integrated data systems in the education sector.

In the face of this demand for ever-increasing scope and interoperability between sector and industry data systems to deliver flexible, seamless, recognized lifelong learning pathways, gaps in existing EMISs have come into stark relief. EMIS are generally understood as the ensemble of operational processes, increasingly supported by digital technology, that enable the collection, aggregation, analysis and use of data and information in education, including for management and administration, planning, policy formulation, and monitoring and evaluation (Law *et al.*, 2018).

> **EMIS are in a dramatic state of evolution. They are increasingly digitalized, integrated, modular and interoperable with other data systems, with the goal of gathering granular, real-time and comparable data on learning, teaching and management processes to help education systems deliver on the right to lifelong learning.**

In many countries, but not all, EMIS data has long been collected, mostly in a conventional manner, including administrative and analogue data collection, processing and reporting. Drawing on multiple sources, such education data generally include information on the number of schools, teachers and learners both in and out of school, teacher qualifications, school performance, expenditures, and inspection outcomes. Many current systems can integrate information about individual learner profiles into education sector administration, including information about age, gender, socio-economic status, additional needs, first language, qualifications, number of years spent in school, and whether learners complete pre-primary, primary and/or secondary schooling. Next-generation systems gather not only this learner data, but also specific learning or transactional data which can be analysed or used to varying degrees.

In face of this promise, however, exists an **asymmetry of information**, with much of the world not having access to timely, quality and complete education data. Globally, **60 per cent** of primary schools, **50 per cent** of lower secondary schools and **35 per cent** of upper secondary schools are not connected to the Internet (UNESCO, 2023*b*). These data make clear that the effective and equitable use of digital data for learning is deeply intertwined with the digital divide. Disconnected learners, as well as marginalized learners and women and girls, risk being under-represented in education sector datasets. The invisibility of certain learners and disproportionate representation of others in these datasets may result in the creation of non-inclusive data-fuelled EdTech, as well as the further marginalization of disconnected communities whose needs remain unseen by policy-makers.

In an increasingly digitalized world, the potential granularity of insights into educational experiences generated from digital data can dramatically expand. However, the continued lack of comprehensive data on the learners in different contexts and settings – including youth and adult refugees, internally displaced populations, migrant learners, and other vulnerable populations such as those with disabilities – can remain a barrier to their educational, economic and social inclusion. Designing EMIS policies, strategies, and governance grounded in equity and inclusion are a pathway to rendering visible the learning pathways and needs of those who remain invisible or under-represented in today's education data systems.

The evolution of education information systems

While the need for digitalized, interoperable and granular data systems in the education sector has long been discussed, actions to meet this need have been slow. Overall, the key objective of this education data system evolution is to improve the efficiency of administration and management. Advanced data systems can incorporate customizable visualization and diagnostic tools to inform policy-makers and the public, thereby increasing the accountability of education systems to deliver on their policy priorities. When connecting aggregate administrative data with process and formative assessment data from LMS, one can "unveil nuances" about educational inequities at the sector level, thus informing planning and management actions in faster feedback cycles (Fischer *et al.,* 2020, p. 132). Such linkages can facilitate processes for school transfer, school choice, university applications, school funding, and central and local resource allocation, among others.

Although the education landscape is increasingly digitalized, over half of countries worldwide continue to rely on paper-based systems that play a purely monitoring and statistical role through the collection and reporting of statistical headcounts, often based on a static school census.

> **60 per cent of primary schools, 50 per cent of lower secondary schools, and 35 per cent of upper secondary schools are *not* connected to the internet.**
> *(UNESCO, 2023b)*

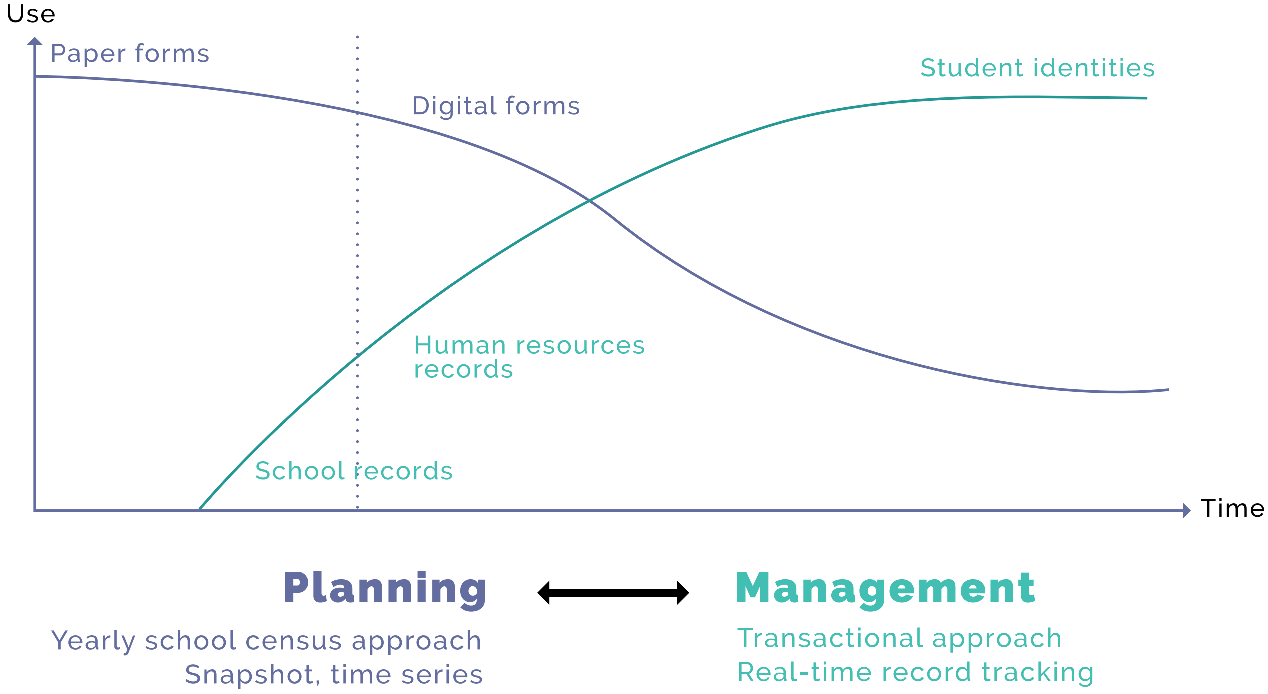
Most countries (54 per cent) have adopted unique identifiers for teachers and students that stay with them throughout their educational journeys, progress is globally unequal, with as low as 22 per cent of countries in Sub-Saharan Africa having unique student identification mechanisms (UNESCO, 2023*b*). Access to real-time management of schools is highly unequal globally, as one in four primary schools does not have electricity and 60 per cent are not connected to the Internet (50 per cent of lower secondary schools and 35 per cent of upper secondary schools do not have Internet connectivity), potential explanations for why 29 per cent of education ministries remain unable to precisely locate schools they govern (UIS, 2020).

These statistics show that countries are at drastically different stages in the process of digitalizing their approaches to information management in the education sector. High-income and low-income contexts face drastically different challenges in advancing EMIS and data for learning. Lower-income contexts may face challenges establishing the necessary infrastructural components to run and maintain EMIS, such as securing sufficient electricity, affording stable and reliable Internet at usable speeds, or procuring school-based devices. They may also lack the capacity – institutional, organizational, financial, technical or human resources – to sustain a digital information management system, leading to reliance on private sector support.

In higher-resources contexts, the challenges may be related to integrating data systems across government sectors, seeking interoperability of data from different sources including non-governmental digital learning providers, improving awareness and transparency around data use in the education sector, and strengthening legislation and standards to protect private data, prevent commercial advertising in educational settings and ensure the ethical use of data-driven practices such as predictive analytics or the use of AI in education.

Despite these varied starting points, the general trends of EMIS transformation over time is visualized in Figure 3.

Figure 3. EMIS transformation over time, from management to planning



Source: Authors.

As illustrated in Figure 3, integrated data platforms for education *management* are at the heart of this changing architecture. However, lower-income contexts are alone in facing challenges to improve the agility of their education data systems. In higher-income contexts with longer histories of digitalization, systems with distinct purposes may have evolved in silos without cross-sectoral governance. Some countries have multiple digital information systems at different stages of evolution, for managing schools (SIS), learning (LMS), teachers and professional development pathways (TMIS), human resources (HR-MIS), technical and vocational education and training (TVET-MIS), higher education (HMIS), and the wider education system (EMIS).

> **One of the major trends in education data infrastructures is to house information management systems on the cloud with interoperability between platforms geared towards learning (LMS) and those with a broader scope geared towards management (EMIS). Enhancing cloud service interoperability can increase the potential for context-sensitive school management and monitoring of policy impact at the school level.**
>
> **However, this architecture also favours the market by lowering the cost of entry into existing digital infrastructures, which in turn can lead to vendor-controlled pricing of data services. It also greatly expands the amount of data that can be collected through the streamlining of data pipelines and single-sign-on capabilities that can collect user data regardless of location**
>
> (See for example Gulson *et al*., 2022)

A strong system-wide strategy is needed to make these systems speak to each other and provide timely, user-friendly insights to education stakeholders. Such strategic governance must be supported by the necessary data and technology standards, quality, coverage, use, competencies, service, software and financing, to make the system run smoothly, safely and sustainably.

Trends in education data infrastructures and architectures

Globally, education data infrastructures and technical architectures are evolving along four major trend lines: digitalization, integration, modularity and interoperability. A new scale of system integration is supported by modern data infrastructures that have shifted data processing from expensive, massive mainframe computers that served individual organizations, to interconnected computers, introducing the concept of cloud service interoperability. The cloud is understood as the central hub for big data processing and analysis, enabling information to be accessed from anywhere. Regarding architecture, traditionally monolithic systems are being replaced with modular system design, with independent modules are designed for specific functions, increasing the overall adaptability of the system, as changes can be made to specific modules without affecting the whole system.

Interoperability is crucial for effective information sharing and collaboration in educational organizations. However, data often remains isolated in separate systems, making it a challenge to access and analyse, and thus hindering the potential uses of data to improve learning, teaching and management. In particular, learning data collected on LMS provided by commercial EdTech platforms may not be interoperable with government systems, because data-related business models restrict data sharing. Therefore, achieving interoperability between government and industry requires considering technical, semantic, organizational and legal aspects.

Technical interoperability involves linking exchange protocols, file formats and services. Semantic interoperability ensures shared information is accurately understood by all parties. Organizational interoperability involves coordinated processes among organizations. Legal interoperability requires aligned legislation to give exchanged data proper legal weight. Overall, enhancing cloud service interoperability between EdTech providers and countries may allow for an even greater shift from centralized management towards a local culture of context-sensitive school management and customized applications tailored to a community's needs (Gulson *et al.,* 2022).

Recently, cloud computing has been eclipsed by edge computing, a method for processing data locally where they are generated. Edge computing can increase the speed and volume of local data analysis by eliminating the need for all data to be sent to a central cloud. Edge data centres generate a wealth of possibilities for locally enabled education actors to understand, manage and solve complex local problems using big data. These increasingly powerful technological advancements – such as cloud to edge technologies, blockchain, fifth generation of wireless mobile telecommunications technology (5G) cellular networks, AI, data warehousing, sophisticated application programming interfaces (APIs) and progressive web apps – are introducing new opportunities for the scale of impact of using data to improve various aspects of education service delivery.

However, these opportunities must be thoughtfully considered against the broader implications they may have on education governance and national sovereignty over education data. Integrated EMIS systems possess immense power to control flows of information and knowledge generation, begging the question, as Zuboff (2019, p. 187) puts it, "Who knows? Who decides? Who decides who decides?"

Though the answers to these questions vary by country, the more education data is unified, centralized and streamlined through the expansion of interoperability infrastructures, the more education governance is diverted onto digital systems that may be grounded in services provided by private companies. As such, governments may find themselves beholden to vendor-controlled pricing of education data services, from platforms to cybersecurity to storage. At any moment, such services that may once have been free may expire and become costed, or the vendor may change their business model and charge for components that are marketed as optional add-ons to basic services, but are in fact essential to be compliant with national data protection regulations.

For example, France and Germany banned the use of Google and Microsoft cloud services in education due to the non-compliance with cross-border data governance, as their cloud services stored data in the United States of America. While the United States and Europe are working to implement the EU-US Data Privacy Framework, announced on 25 March 2022, which would ensure that all data transfers between the United States and Europe are General Data Protection Regulation (GDPR)-compliant, **Microsoft** released a product called "Microsoft Cloud for Sovereignty" to help customers leverage the cloud while simplifying the architecture of ensuring regulatory compliance, deploying sovereign controls and maintaining data residency (Sanders, 2022). With such large-scale products, the global government cloud market is expected to nearly triple in six years according to market research, rising from US$27.6 billion in 2021 to US$71.2 billion by 2027, with the first national customers coming from Europe (Mukherjee, 2022).

The cost of such cloud-based and interoperable data system architectures goes beyond ensuring compliant data governance. Integrating advanced technologies such as the generative AI tool ChatGPT into education services comes with significant computing costs. As a result, EdTech providers are increasing the costs of their services if such sophisticated data analytic power is to be unlocked at the classroom level. For example, **Khan Academy**, an American non-profit educational organization whose set of online educational tools – including basic analytics, online lessons and other school services – are used by over 20 million learners in over 190 countries, charges American school districts an annual fee of US$10 per student. However, in the upcoming school year, those schools that would like to pilot test their new intelligent tutoring service powered by ChatGPT, called Khanmigo, will see their annual fee per student increase by 600 per cent, to US$60 per student (Singer, 2023).

As mentioned above, the potential to integrate such tools into broader education management architectures is unevenly realized around the world. Many countries do not yet have the connectivity or technical infrastructure required to incorporate the most dynamic, powerful elements of advanced technologies into their education data systems. In many countries, the lack of tangible and exact data on network coverage and connectivity, especially in peripheral and agricultural areas, is in itself a barrier to expanding such data systems and digital education opportunities. Telecommunication companies, such as WGD4L members **Millicom** and **Ericsson**, work to bridge these rural-urban divides by identifying and connecting communities to enable the potential benefits of data-*informed* education. Even high-income countries face challenges with their data system architecture, for example in moving from cloud to edge computing, or expanding existing digital infrastructure to remote communities.

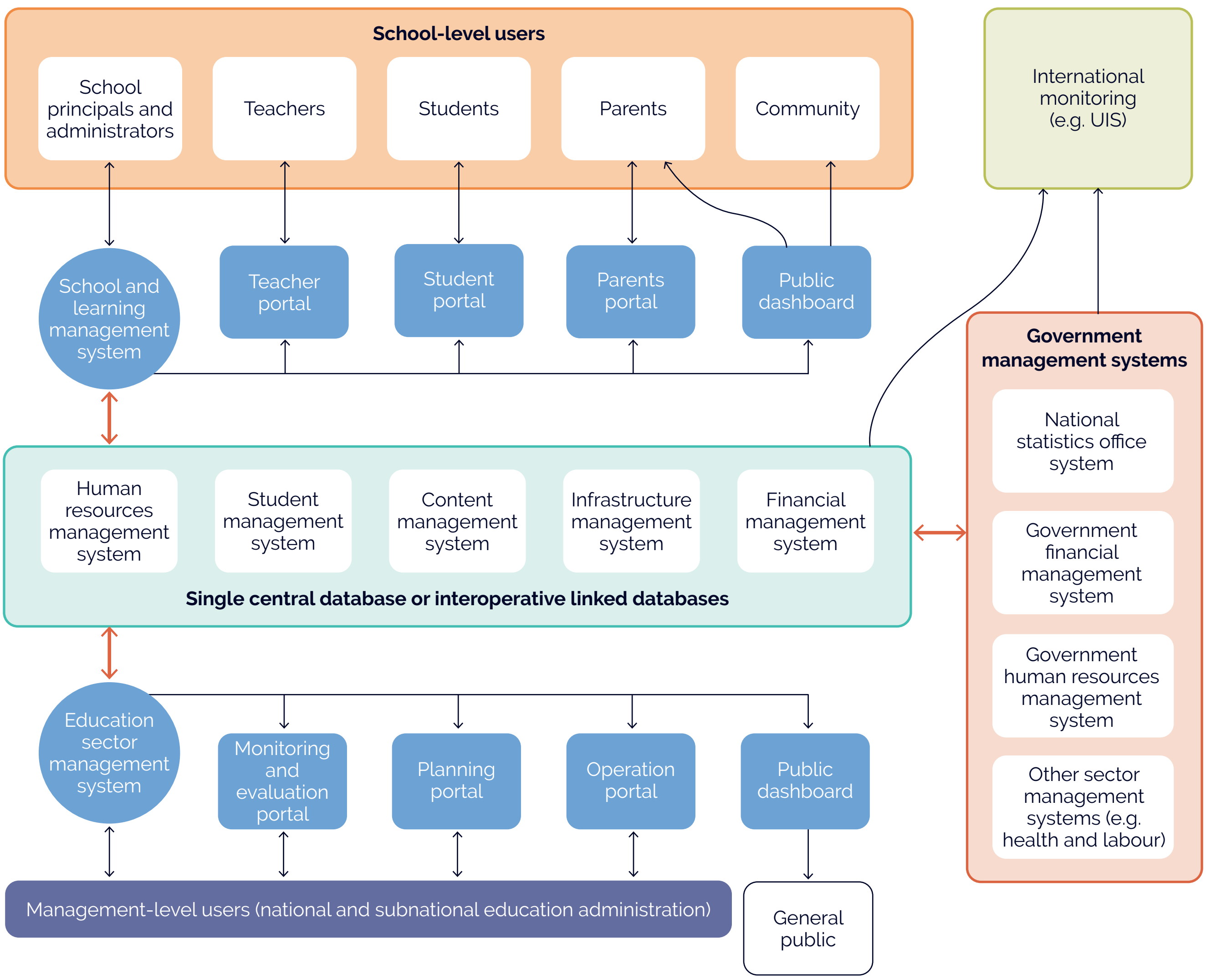
Where lower-resource countries decide to move towards integrated, modular, open data ecosystems, significant investments will be needed to update their data architecture and cover the costs of data-migration projects, and ecosystem safety and maintenance. These investments are not insignificant. The *2023 Global Education Monitoring Report* (UNESCO, 2023*b*) on technology and education estimates that a full digital transformation of education with Internet connectivity in schools and homes would cost over **US$1 billion per day** just to operate.

The link between data for administration and data for learning

Investing in interoperable, user-friendly data architectures at the administrative level can influence the potential transformative uses of data for learning beyond sector administration and management. Access to actionable, timely information services for all levels, spaces, and stakeholders of education can benefit a range of factors, including learning, teaching, teacher management, quality assurance, accountability, lifelong learning, inclusion and resilience. These links between data for learning and data for management are visualized in Figures 4, 5 and 6 below, which provide abstracted diagrams of data for learning system architectures.

Figure 4 depicts **UNESCO's** vision for a sector-wide education data system where EMIS functions are horizontally and vertically integrated to create real-time public dashboards, and connecting with other sectors' information systems, such as health, finance, labour and social security. The future of EMIS could be one where powerful analytics support real-time and data-*driven* decision-making at all levels of education, from day-to-day operations to strategic planning functions.

Figure 4. UNESCO vision of EMIS vertical and horizontal integration



Source: UNESCO (2022c).

Figure 4 depicts an EMIS where granular information is integrated across the micro-, meso-, macro- and global levels both vertically and horizontally. Practically speaking, vertical integration means a system that supports secure information flows from the level of learning (i.e. learner performance and digital learning data) to school operations (daily school activities and attendance data), to local management and control (resource distribution data for technical operations), to planning and strategy (central government general strategy and education policy), to national or global reporting (SDG 4 or other national/global targets).

Horizontal integration is an EMIS that is linked with both sector-wide and government architectures. This could mean a ministry of education information system connected to the ministry of higher education or other tertiary information systems, and also to other ministries directly administering education services, such as the ministry of labour or health. For wider horizontal integration, the EMIS would also be linked with the wider government data architecture, including the civil registration bureau, the ministry of public administration or finance, or the national security organization.

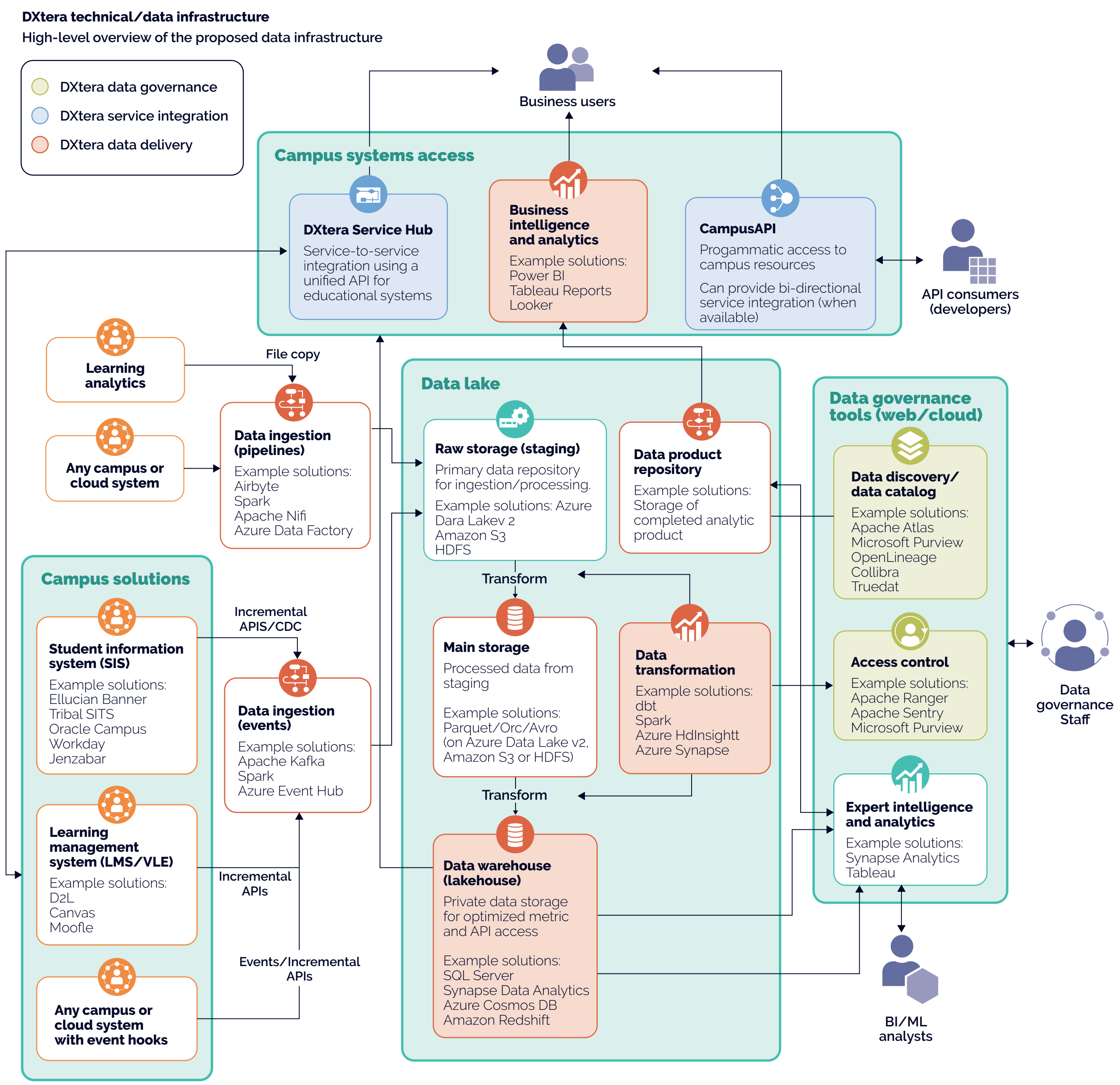
Overall, UNESCO's vision for a future-ready EMIS is one of integration between EMIS and LMS through real-time learning data flows secured with protective measures such as anonymization and encryption to allow for more responsive, relevant and contextualized management of education sector resources, and monitoring of policy impact. To safely achieve such levels of integration, it is necessary to have strong cross-sectoral and trusted multi-stakeholder collaboration within a closely monitored, securely controlled and legally compliant model of public governance. UNESCO supported the transformation of this vision into an operational initiative called OpenEMIS, which is structured to assist lower-income countries to develop and maintain a contextualized, modular national EMIS within a lifelong learning perspective and built through open-source software. By offering capacity-building and technical support for smooth operation and administration, the OpenEMIS Initiative presents tailored solutions to address capacity limitations and ensure effective implementation.

Another example of a vision of integrated data systems comes from the United States, where the **DXtera Institute** is working to build trust between education entities and technology solutions providers in order to better exchange, utilize and optimize data from disparate systems or applications. The DXtera Institute is a non-profit membership consortium of education entities, and firms which develop shared technology solutions and practices that address the complex barriers and technology challenges regarding infrastructure, data management and integration by developing open and modular technology solutions.

Figure 5 showcases a cloud-first self-service deployment of DXtera's EduMesh data management platform. Its primary focus is to streamline data acquisition and management processes through a data lakehouse architecture. This solution enables the integration of institutional data through a comprehensive set of educational domain models initially developed at the Massachusetts Institute of Technology.

The EduMesh platform in Figure 5 illustrates a standard set of interfaces and models for interacting with data, enabling institutions to communicate and share their reports, analytics and tools across the entire DXtera member base. Therefore, it seeks to facilitate the development of common tools, reports and analytics to address shared challenges encountered by multiple institutions. All components of the EduMesh platform are open-source and modular, designed to be interconnected to meet diverse educational business needs at the institutional or system level. The depicted infrastructure goes beyond reporting and descriptive analytics. The platform also offers comprehensive and openly defined software APIs that enable integration across different business domains and legacy systems, providing data and system access for analytics, reporting and application integration. These components empower institutions to maintain and control data in real time across their systems of record.

Figure 5. DXtera Institute technical/data infrastructure



Source: DXtera Institute (2023).

Although focused on higher education, Figure 5 visualized a data system architecture applicable to any educational setting or data decision level (local, regional, state or national). These complex diagrams depict the growing trend of merging administrative management with learning management, or in other words the meshing of EMIS and LMS through cloud interoperability. They demonstrate the evolving architectures of education systems to become a more granular, integrated system with powerful analytics that supports real-time and data-*driven* decision-making at all levels of education, from day-to-day operations to strategic planning functions. However, the aforementioned risks related to market profit from such expanded interoperability must be considered in order to ensure transparency, trust and accountability between institution and vendors regarding the short- and long-term costs of sustaining seamless, modular and granular data systems, and how to ensure that multi-stakeholder data system partnerships are designed with privacy and ethics as grounding pillars.

To further elaborate this trend, in a Working Group meeting on 25 April 2023, the **OECD** identified four kinds of next-generation information management systems, which can link and integrate administrative data with learning data:

→ **Reporting and research data systems**
Some of the most advanced student information systems so far, such as the Ontario School Information System (OnSIS) in Canada and the Sistema Integral de Resultados de las Evaluaciones (SIRE) in Mexico have been built as reporting and research data systems. Using individual-level data, they collect and present statistical reporting and evaluation with a traditional focus on the accountability of the systems and the educational institution's performance. The main use of this kind of system is to provide information to policy-makers and the public, but, in some cases, these systems have also been designed to develop research capacity about educational issues.

→ **e-Government data systems**
Some systems have been inspired by the e-government approach promoting automated data integration across government agencies, such as the Estonian Education Information System (EHIS) and the National Education Information System (NEIS) in the Republic of Korea. Digital identities or ID cards facilitate, for example, the tracking of learners as they move from one educational institution to the next, enabling more informed decisions on resource allocation. In some cases, these types of data systems are not designed to track student performance, but to improve administrative solutions. They also offer the potential for creating links with data from other sectors.

→ **School improvement data systems**
Another type of system identified by OECD are those designed to support efforts to improve an educational institution, including England's Analyse School Performance (ASP) system and Portugal's Escola 360° (E360°). They achieve this by putting key data, via features including customizable school reports, visualizations and dashboards, into the hands of principals and teachers. These systems are also typically used by regulators before they visit an institution. They can also facilitate digital communities of practice and enable new improvement routines such as data teams.

→ **Expert data systems**
Expert data systems such as the statewide Colorado longitudinal system and the SchoolView website in the United States incorporate something of all the preceding three systems. They offer real-time information and feedback to learners, teachers, and administrators, facilitating continuous and ongoing adjustment to learning pathways as opposed to stagnant end-of-cycle feedback. This is achieved through a combination of administrative data with process and formative assessment data from LMS. Expert data systems can also link to banks of educational resources, recommendations and networking platforms for teachers.

Expert data systems are able to offer more actionable information across multiple aspects of learner and learning management with the technology now existing, to go beyond mere reporting and research capabilities. EdTech companies increasingly offer such expert data system services, such as **Anthology's Blackboard Learn**, an LMS that is widely used around the world by both educational institutions and industry. Their data for learning model provides online and blended learning experiences that can integrate with existing data ecosystems. As depicted in Figure 6, this integration ensures that data is being shared, and can therefore create better learning experiences.

Figure 6. Blackboard Learn layer model

	Layer	Audience	Expectations
Experience: Personalized; Contextual; Useful; Actionable	**Intelligence**	Students, instructors, advisors, academic staff	Recommendations, suggestions, alerts
	Context	Students, instructors, advisors, academic staff	Comparison to better contextualize information
	Information	Students, instructors, advisors, academic staff	Present raw/summarized data for more informed decision-making
Data	**Reporting**	Analysts, institutional researchers, leadership	Querying, dashboards, custom reports
	Access	Developers, data scientists	Raw data access, REST APIs
	Harmonization	Developers, data scientists	Bring data from silos together Data integrity

Source: Presentation to WGD4L by Nicolaas Matthijs, Vice-President Product Management, Anthology on 23 May 2023.

The model comprises six layers, split across two superlayers. Anthology refers to the first as the **data superlayer**; within that, there are three sublayers. The technical nature of the data superlayer means it can be difficult for educational institutions to fully handle independently, and thus Anthology provides system technical support. The first sublayer is the **harmonization layer**, which is necessary for educational institutions that may have a large number of internal and external silos, typically corresponding to the different solutions in use. The harmonization layer ensures that these different data sources can be brought together in a way that ensures data integrity across them all.

Above harmonization is the **access layer**, which offers educational institutions technical access, including the use of APIs, to the underlying raw data after it has been harmonized. The final sublayer in the data superlayer is the **reporting layer**, which offers more day-to-day access to the data via more commonly understood data dashboards, often available to institutional leaders and researchers.

Above the data superlayer sits the **experience superlayer**, which is where the data and insights from the information start to show up directly within the user experience, within the product, and directly to instructors, learners, administrators, etc. The first sublayer here, the **information layer**, offers students, instructors, advisors and other academic staff access to raw or summarized data within the context in which it is useful, enabling them to make informed decisions. The **contextual layer** follows, comprising raw, summarized data, which are able to provide context and allows for comparison, for example between learners in a class. The top layer is called the **intelligence layer**, which uses data to inform intelligent actions such as making recommendations and automating specific actions.

The middle layers of Blackboard Learn can provide an instructor with raw information to help them make informed decisions in a way that is actionable. For example, the information layer can let an instructor know if a student has not completed an assignment and send them a reminder. The contextual layer can show student data next to each other within the context of the cohort or class. Unlike consumer markets, the education sector typically has a lot of unstructured data that lives in a variety of systems, which means that the intelligence layer – which can include systems for predictive analytics and other intelligent analysis – must be used with caution and transparency, in contextually appropriate situations.

The emergence of generative AI brings much opportunity to future expansion of the intelligence layer, as generative AI can process unstructured data and generate high-quality content when prompted with high-quality queries. Given the varied needs between different education data system levels of maturity, there remains an opportunity for continued improvement across all six of these layers. The ultimate goal of this LMS architecture is to improve the learner experience and make course instructors, advisors and others more efficient and effective by gaining deeper insight into the data.

Box 3. Blackboard Learn: Lessons learned at Keiser University

For Keiser University in the United States of America and each of its 21 physical campuses, it was essential to deploy a learning management solution university-wide, while also offering the ability for each campus to personalize according to their requirements. The university took advantage of the customization tools available within Blackboard Learn, ensuring that each campus could meet the needs of students and their specific learning environments. The ability to acquire comprehensive visibility into educational performance and trends was crucial for driving continuous improvement through data-supported decisions. According to Blackboard, this project has enabled the university to consolidate historical data, gaining a holistic understanding of student journeys and providing more effective guidance.

The data layer has empowered Keiser's staff to gather information from different sources and create intuitive dashboards, enabling leadership and faculty members to quickly identify at-risk students and intervene, leading to higher retention rates. Within the experience layer, faculty members and assessment teams now enjoy easy access to assessment data for programmes and courses, implementing a robust review cycle that fosters increased student engagement and ensures an enriching learning experience, ultimately contributing to the success of their students and the continued growth of the institution. By allowing each campus to tailor its approach without compromising university-wide collaborations, Blackboard Learn has empowered the institution to continually innovate and make data-*driven* decisions regarding its practices and procedures (Blackboard, 2022).

The potential of integrated, interoperable education data systems: Uses and users

As evidenced by the models and cases explored above, the growing trend of interoperable LMS and EMIS data architectures opens new possibilities for informed decision-making for education stakeholders throughout the system. Interoperable education data systems can ensure the alignment and harmonization of national, subnational and school-level strategic plans and policy actions, promoting coherence and accountability within the education sector. By linking these plans, the system fosters a unified approach to achieving educational goals, streamlines reporting processes, and facilitates effective monitoring and evaluation. However, such architecture also raises concerns regarding corporate control of education data flows that data governance decision-makers must understand and safeguard against.

Each of the subsections below outlines some of these possibilities afforded by integrated data systems, organized by different user profiles, from learners to school inspectors to central managers, given their varied entry points to the data system. Illustrative use cases of interoperable LMS and EMIS around the world provide concrete cases of these possibilities put into practice.

Lifelong learners

The broadened vision of a lifelong learning journey, beginning in early childhood and continuing through adult education, has been accompanied by an expansion of data systems that integrate comparable and longitudinal data on an individual's learning outcomes from various subsectors over time. For example, a complex subsector like early childhood care and education or TVET can benefit from integrating data from multiple sources, such as household surveys, labour-market data or microcredentials acquired during short courses or trainings, to provide a more holistic picture of learner needs and circumstances (UNESCO, 2023*a*). The Global Farmer Field Schools (FSS) platform of the **Food and Agriculture Organization of the United Nations** (FAO) is an example of how integrating learning data from multiple sources into one community platform can encourage knowledge sharing to promote lifelong learning for farmers, livestock herders or fisherfolk on how to shift towards more sustainable production practices.

Like other dimensions of data for learning, data system integration and interoperability can be seen as a double-edged sword for lifelong learners. On the one hand, ensuring that learning data is aggregated across periods and sources can be used to improve government services that respond to child vulnerability to harm and deprivation, such as asylum-seeking children or undocumented youth (Livingstone and Pothong, 2022). On the other hand, such integrated systems that provide a seamless transfer of comparable records between institutions could spur unnecessary surveillance mechanisms by tracking years of granular student data that may make negative data points (behavioural or academic) from which it is difficult to recover, thereby influencing negative institutional and self-perception of their success in learning systems.

Expanding the right to education beyond the K-12 and into adult learning and lifelong learning could have a positive effect on promoting continuous learning in the context of rapidly changing workforce requirements. Alongside the recognition of lifelong learning as a right to which all people are entitled is the possibility of choice. Learners should be able to opt for different types of digital learning opportunities and providers, which could include private and public options, depending on their needs, preferences and modalities. Integrated platforms can support such freedom of choice by compiling and linking to various learning options in one digital space.

Teachers

Teachers directly collect more learning and learner data than any other actor in the education system. They are often responsible for gathering and submitting administrative data, such as attendance or family contact information, academic data such as coursework and formative and summative assessments, and behavioural data such as misdemeanours or positive behaviour reinforcement mechanisms. Some particular teacher-facing benefits of increasingly interoperable systems could include illuminating student learning or attendance trends, enabling comparative analysis between classes or schools, sharing lesson plans or tools within teacher communities of practice, tracking professional development and career performance, and improving communication with students' families.

Early school leaving remains a major barrier to learning equity around the world. According to UIS, more than 64 million primary school students dropped out of their education in 2020, with higher rates in low- and middle-income countries.

To counter this, teachers may benefit from sophisticated data systems that can alert them when a learner's data displays early warning signs for course failure or drop out. They can then combine such predictive analytics with their own qualitative and contextualized insights to engage and support students or their families, reducing the likelihood of them leaving the educational system prematurely.

Interoperable school or LMS that have entry portals for parents make it easier for teachers to engage with guardians by communicating school announcements or information about attendance, learning progress, behavioural concerns and more.

LMS with real-time and individual-level progress monitoring capabilities can also support inclusive education by assisting the early identifying of children with disabilities, and supporting teachers to determine what additional support may be needed for individual learners. If such systems are interoperable with the centralized EMIS, teachers and school leaders can support evaluation of the special needs services provided, the number of specialized support staff or special educators within and between schools, and the degree to which services are provided equitably across the system. For example, in the Republic of Korea during COVID-19 school closures, inclusive data systems ensured that all students with disabilities were individually assessed and provided with customized learning materials, including digital content with subtitles and/or sign language, materials in Braille, assistive devices, and even home visits to track the efficacy of these interventions (UNESCO, 2023*b*).

Box 4. Supporting individual learning needs through interoperable data systems in Saudi Arabia

In **Saudi Arabia**, the Unified Digital File programme aims to develop a consolidated digital system that renders all student information and data interoperable between EMIS and LMS platforms in order to promote transparency and support future decision-making in education. Digital files for students include personal information, psychological, social and educational data, and a record of skills and knowledge to measure learning outcomes and to facilitate the use of diagnostic tools, including the detection of students at risk. The LMS Madrasati ("My School") works as a comprehensive e-LMS and is linked to a national information system for students. Assigned digital content is aligned with learning goals and performance-monitoring dashboards. It is equipped with educational tools that cover areas such as scheduling, learning objectives, virtual classrooms, an enrichment resources bank, and e-courses, learning paths, digital content and an eLearning dashboard. This integrated profile and assessment system includes several initiatives geared towards the inclusion of students with special needs, ensuring that gifted students and those with disabilities receive the support they need through an assessment system of all areas. This assessment system sits within a single integrated digital system to enable comprehensive student assessment tools that integrate student records that measure learning, behavioural and skill progress at all education levels.

> **According to a study in the *Global Education Monitoring Report 2017/8*, 56 per cent of teachers surveyed in England argued that data collection and management caused unnecessary workload for them. In Finland and Sweden, 75 per cent and 95 per cent of teachers respectively reported an increase in demands on their data collection, and nearly all teachers and school leaders surveyed viewed workload as a serious problem.**
> *(UNESCO, 2017)*

Integrated and modular data systems may also enable teachers to access high-quality digital educational resources stored in digital content repositories. A teacher-facing module within a larger EMIS could allow teachers to create digital resources, online courses, assessments or reporting dashboards, and share these resources with other teachers in a systematic way. Such user-friendly data systems can build individual teacher agency, as well as communities of practice where teachers support other teachers by sharing tools and experiences.

However, while data systems can empower teachers to differentiate and improve their inclusive pedagogical practice, it can also carry a heavy administrative workload. Integrating new data-driven digital tools flexibly and fluidly into teaching practice can be challenging and can divert attention from teaching and towards data collection for the sake of it. The *2017/8 Global Education Monitoring Report* (UNESCO, 2017) on accountability in education illustrated that the limited training on data processes, combined with the pressures to continuously collect data for accountability's sake, had resulted in teacher frustration and a decline in the attractiveness of the teaching profession. For example, 56 per cent of teachers surveyed in England argued that data collection and management caused unnecessary workload for them. In Finland and Sweden, 75 per cent and 95 per cent of teachers respectively reported an increase in demands on their data collection, and 93 per cent of teachers and some school leaders viewed workload as a "very" or "fairly" serious problem (GEM Report, 2022). As the *2023 Global Education Monitoring Report* (UNESCO, 2023*b*, p. 21) recommends, technology in education should put learners and teachers at the centre to prevent it becoming a burden or an impediment to teachers' ability to teach.

(UNESCO, 2017)

Box 5. Predictive analytics to counter early school leaving in the United Kingdom of Great Britain and Northern Ireland

The Open University, the largest distance education provider in the United Kingdom, is engaging in long-term Predictive Learning Analytics (PLA) through its system OUAnalyse, which aims to support teachers to monitor the progress of students participating in distance education. Data employed to predict students at risk include demographic information, pre-course results, assessment data as well as data from their virtual learning environment. The system allows teachers to act before students fail by systemizing the monitoring of what students were doing at a certain time. The system also explains why a certain at-risk prediction is made, allowing teachers to correct the prediction. Based on data from a four-year longitudinal study, during which 1,182 teachers accessed OUAnalyse with the system, reaching 23,630 students, teachers perceived the tool as a means to identify students who were struggling who needed extra support. It is also worth highlighting, however, that engagement with the tool varied, with the researchers calling for further research into the tool (Herodotou *et al.*, 2020).

Teacher trainers

A human resources module or subsystem of an EMIS can be used to support all areas of teacher management at the level of each individual teacher, in particular professional development, performance management and teacher deployment. Such systems can recommend continuous professional development courses based on teacher interests and growth areas, informed by data from classroom observations of teaching practices, self-evaluation tools, formative and summative student assessment data, teacher assessments and appraisals, and data generated through an LMS.

For example, in England, the EMIS staff performance module manages all the information related to the evaluation and performance of personnel. It also guides the creation of personal goals, such as the management of professional development, in-service training, teacher evaluation from the standpoint of student performance, and development and career plans. In a system that can collect longitudinal data, teacher performance and continuous professional development can be monitored over time and transferred between institutions, thereby improving opportunities for teacher mobility and professional growth. From the administrative perspective, EMIS can be used to automatically verify the identity of all education personnel, thereby preventing the payment of ghost educators, and can be used to partly automatize teacher allocation processes following pre-defined rules.

School leaders

Integrated education data systems offer school leaders valuable data-*driven* tools to reduce rates of out-of-school children or early school leavers, in collaboration with teachers. EMIS can help school leaders identify out-of-school children and develop plans to support their enrolment in school. The interoperability between LMS and EMIS through unique learner IDs allows for better coordination and intervention strategies, including monitoring absenteeism, facilitates targeted actions to support students and prevent dropouts.

Overall, school leaders are the linchpin between centralized policies and school-level practices. Therefore, their access to – and reporting of – real-time data from both the micro- and macro-level education data systems allows for seamless top-down and bottom-up data flows. This not only provides central managers and policy-makers with the ability to monitor the implementation and impact of education policies, but also enables school leaders themselves to inform their own management of teaching, learning and operations through easy access to comparative analysis between schools with similar profiles.

> **Worldwide, an estimated 244 million children and youth are out of school. Dropout rates have risen and are highest among older children. Integrated, comprehensive education data systems can help school leaders identify these children and develop strategies to bring them to school.**

School inspectors

Intelligent warning systems and access to comparative data between schools, districts and regions can make modern EMIS architectures an asset to school inspectors. Comprehensive information on the quality of service delivery within each individual school – including teaching and learning process, access to inclusive education services, and the overall sanitary and environmental conditions of the school – can empower school inspectors to make informed decisions, and drives continuous improvement in the education sector through improved accountability.

School buildings are responsible for providing numerous services beyond teaching and learning, including clean water, nutritious food, spaces for physical movement, and access to safe and meaningful Internet connectivity, among others. An EMIS can play a crucial role in identifying schools in need of immediate supervisory support, for both operational and teaching and learning purposes. Through pre-defined indicators of weak performance, the system automatically flags schools that should be prioritized for supervisory visits, reducing risk by ensuring timely interventions are implemented to address their specific challenges.

An EMIS supports the management of school support services, including school health, nutrition, psychosocial support, special needs education services, and interventions. To enhance transparency and accountability, automated school and district report cards can be generated, containing vital information on learning outcomes, teaching practices, inspection findings, and quality, equity and access indicators. **Brazil**, for example, has developed a synthetic indicator called the Primary Education Development Index (IDEB) to evaluate educational quality at the school level, with values between 0 and 100. Moreover, Brazil's EMIS is integrated with health data, allowing for the identification of learners' health issues, which can be consulted on the same platform, and thus inform a broader understanding of the well-being of the school community (Arias Ortiz *et al.,* 2021).

Digitized, real-time EMIS can support school inspectors in forming a comprehensive view of the quality of service delivery of schools by creating comparative data dashboards between schools that use key indicators, including teacher retention, absenteeism and national assessment performance. Such systems can offer insights into the performance of individuals such as school leaders and sector managers, as well as educational levels and departments. In **England**, the ASP system of the Student Information Management Software (SIMS) SchoolView module offers a complete overview of schools in terms of the key performance indicators established, such as student attendance, number of students, conduct, evaluation, information about personnel and the status of students with special educational needs.[3]

When school inspectors have access to powerfully interoperable data systems, they are empowered to ensure that national, subnational and school-level strategic plans are in alignment by tracking progress against set targets specified in strategic plans. This information then can be passed up towards central managers who can further monitor higher-level administrative areas such as districts and regions. In **Pakistan**, for example, EMIS data are openly accessible through an interactive map and dashboard on a central government website,[4] which sits within the federal Ministry of Education and is interoperable with provincial EMIS, enabling the creation of district reports and the identification of low-performing schools.

3 See https://www.ess-sims.co.uk/products/sims-schoolview.

4 See http://www.emis.gob.pk/website/Default.aspx.

Parents and caregivers

Parents can actively engage in their child's education through education data systems that provide secure insights into student learning, evaluation of school quality, and streamlined communication with teachers. An integrated EMIS with a parent module can give parents a better understanding of their child's progress, identify areas that may require additional attention, and actively engage in their educational journey. This information can empower parents to make informed choices regarding their child's education by considering factors such as school performance, school resources and curriculum, and the overall quality of educational opportunities. In contexts where school choice is an option, parents can ensure that they select a school that provides the learning environment they seek for their child.

Access to learning data can help parents support their child's learning journeys by understanding where they may need additional support. They can also provide platforms for agile communication between teachers and students regarding their child's experiences at school. As such, these data systems can strengthen the partnership between parents, caregivers and educational institutions, fostering an environment that promotes student success and well-rounded development. For example, in **Estonia**, local authority schools in Tallinn use data from the census and register of households for registration in first-year primary education. Based on this data, children of the age to start primary school are identified and parents are contacted by email to confirm enrolment and enquire about their preferences regarding the school. Subsequently, the data sent by parents is updated in the EMIS, the children are assigned to a school, and the parents are notified of the result (OECD, 2021*a*).

The security and privacy of sensitive information on learning journeys and learner identities is a prime concern for parents, caregivers and students themselves. Today, the integration of blockchain technology within EMIS architectures can ensure the creation of portable, verifiable, and resilient learning records and credentials for students. This innovation enhances international mobility by allowing educational achievements to be securely stored and easily accessed across different institutions and jurisdictions. The use of blockchain technology provides the necessary security and privacy measures to protect students' data, ensuring the confidentiality and integrity of their educational records and empowering their ownership of data.

Overall, parents should be made aware of how their students' data is aggregated from different sources. However, a recent report from the Digital Futures Commission argues that currently in the **United Kingdom of Great Britain and Northern Ireland**, local authorities are not transparent about how they use and link data about students and their families, causing high levels of distrust in parents from marginalized social groups who were suspicious that their personal data – such as information regarding support from social services, divorce status or financial situations – could be misrepresented, misjudged or misused through biased labelling (Livingstone and Pothong, 2022).

Community and civil society

Education data systems that provide simplified school profile cards allow communities to evaluate the quality of service delivery in schools. These profile cards provide transparent data on various aspects of school performance, including achievement of performance indicators and the utilization of school funds. By accessing this information, community members can actively participate in monitoring school improvement plans, holding educational institutions accountable and advocating for necessary changes to enhance the quality of education.

Through information services, grievance mechanisms, and access to data for policy dialogue, communities can evaluate school performance, monitor improvements, voice their concerns, and advocate for positive change. These data systems foster transparency, accountability, and collaboration between stakeholders, ultimately contributing to the enhancement of educational outcomes and the well-being of the community as a whole.

Uruguay, for example, has advanced approaches to sharing information about the performance of the education system to different civil society actors to encourage oversight of institutions. The Council for Early and Primary Education (CEIP) discloses information on institutional performance to the general public through an educational monitor and GURÍ Familia, a digital platform under the direction of the Division of Information and Technology within the Directorate-General of Early and Primary Education, gives parents the ability to easily monitor students' performance (Arias Ortiz *et al.,* 2021).

With comprehensive information on educational indicators, outcomes, and trends, civil society organizations can engage in informed discussions and advocacy efforts. By leveraging data, these organizations can effectively contribute to policy development, implementation and evaluation, ensuring that educational policies align with the needs and aspirations of the community. The availability of reliable data strengthens civil society's capacity to influence decision-making processes, and fosters collaboration between stakeholders for sustainable improvements in the education sector.

Policy-makers

Integrated, interoperable, secure EMIS can empower education sector managers by providing them with comprehensive information and tools for effective decision-making and management. By digitizing processes, accessing timely information, tracking key indicators and managing support services, sector managers can drive improvements in educational outcomes, allocate resources efficiently and make informed policy decisions. These data systems enhance the overall governance and management of the education sector, leading to positive impacts on teaching and learning.

One of the major incentives to invest in a comprehensive EMIS architecture is to digitize, automatize and improve various management and administrative processes in the education sector, including those related to the management of schools, students, personnel, infrastructure, and the allocation and tracking of educational materials such as textbooks and teaching materials. Investment in the digitalization of such processes then engenders more data production, and as a result, more potential to generate valuable insights from this data to make sector management more efficient, transparent, proactive and reactive in the face of new and anticipated sector challenges.

Beginning in 2018, the federal Government of **Nigeria** has been investing in the development of a tool for the collection of timely, credible and reliable data that is compatible with mobile devices. This means that data can be retrieved from even the most remote school site anywhere in Nigeria, instead of stakeholders waiting for years before data collection and reporting processes are completed. With the help of technology, the Ministry of Education has been able to establish a robust data centre with the capacity to hold data from all levels of education in Nigeria (Group of 20 [G20] Education Working Group, 2023).

Box 6. Sierra Leone's digital school census

In a strong cross-sectoral collaboration with the Ministry of Finance, the Ministry of Education of Sierra Leone successfully transitioned to a digital school census in 2019. This transition enabled evidence-based financial allocation decisions as part of the Free Quality School Education (FQSE) programme for the most underserved communities, based on the accurate collection of enrolment and infrastructure data for all 11,000 primary schools in the country (including pictures, GPS coordinates, data on absenteeism and a teacher database). The conversion of the annual school census form to an open data kit format, and the procurement of solar-powered tablets, were key to the digital transition. Rapid data visualization formats are further helping to inform decision-making at the district and national levels (Namit and Thi Mai, 2019).

By providing disaggregated, longitudinal information through dashboards and reports, EMIS can help sector managers identify trends over time in a range of areas, including budgetary expenditures, comparisons of learning outcomes and teacher competencies between regions and districts, availability of instructional coaching and leadership at schools, school connectivity, and resource distribution, among others. For example, EMIS tools may support the linking of plans at both national and subnational levels, and the linkages of these plans to budgets, such as total investment by school and the according academic outcomes. This information facilitates evidence-based decision-making and empowers managers to take targeted actions to prioritize areas for improvement within the education sector.

Promising practice: Towards an integrated education information ecosystem in India

At the policy level, India, the 2023 President of the G20, has made major advances on improving integration, digitalization, interoperability and modularity in their central and state government education information systems through its Vidya Samiksha Kendra (VSK) and planned Education Ecosystem Registry (EER).

The VSK is a series of public dashboard intended to break information silos and help the Ministry of Education effectively leverage its multiple initiatives. It provides all education stakeholders with the ability to track progress of initiatives, make sense of data and coordinate improvement efforts based on data, or "see, act and solve" (National Digital Education Architecture, 2021). These anonymized big data can guide better delivery of student services, innovation and research by EdTech companies and the rest of the ecosystem to facilitate the development of personalized learning solutions. All open data from the registries and transaction systems is anonymized when transferred to VSK systems, which enable data-*driven* observability and decision-making at the policy decision-making level.

As next steps towards the data-*driven* management of the education sector, India is planning the EER, which seeks to establish a national-level bird's eye view of the country's student and teacher cohorts. The EER will be synchronized with the Unified District Information System for Education Plus (UDISE+), which is one of the largest management information systems (MIS) initiated by Department of School Education and Literacy, covering more than 1.48 million schools, 9.5 million teachers and 265 million children (G20 Education Working Group, 2023 [forthcoming]). The EER is expected to be an integral part of the National Digital Education Architecture (NDEAR) and will form part of the national identity frameworks.

The entire system is online and has been collecting data in real time since 2018/19. UDISE+ has a mandate of collecting information from all recognized schools imparting formal education, from pre-primary to class 12. Similarly, to portray the status of higher education in the country, the Ministry of Education conducts an annual web-based All India Survey on Higher Education. The survey covers all the institutions in the country engaged in imparting higher education. Data is being collected on several parameters such as teachers, student enrolment, programmes, examination results, education finance and infrastructure. In so doing, EER helps users to identify personal learning and growth needs. Indicators of educational development such as institution density, gross enrolment ratio, pupil-teacher ratio, gender parity index and per student expenditure are also calculated from the data collected through the All India Survey on Higher Education to help make informed policy decisions and research for the development of the education sector.

The EER and VSK are compliant with the NDEAR, which was launched by the Prime Minister of India in July 2021 with a vision to catalyse the education ecosystem. The NDEAR Ecosystem Policy, published publicly in November 2022, is a major engine driving the fulfilment of the National Education Policy 2020, which aims to encourage the development of innovative, inclusive and contextual solutions to meet the needs of all teachers and learners around the nation.

Promising practice: Investing in data for learning as a priority in South Africa

As part of the National Development Plan 2030, the Department of Basic Education implemented the South African School Administration and

Management System (SA-SAMS). The out-of-the-box open EMIS has been designed for use across the South African education sector as a unified open data platform to support standardized education policy implementation across the nation and all regions. SA-SAMS sits at the heart of South Africa's education modernization efforts, supporting tactical operations at the district level and strategic actions at the provincial level, while also informing policy decisions at the Department of Basic Education.

As well as improving operational administration and management at the school level, the mandatory reporting system also collates learner, teacher, management and administration data to support operations across the entire education sector. SA-SAMS provides real-time, validated data to help deliver improved data-*driven* educational decisions. The system helps over 10,000 schools across South Africa manage and administer systems including human resources, learner and parent information, governance, curriculum data, timetabling and more. The Department of Basic Education reports that SA-SAMS has more than 15,000 daily users, including principals, educators, administrative staff and Department of Basic Education-supported staff (South African Department of Basic Education, 2022).

A recent assessment of the SA-SAMS model defined and documented the benefits that would be realized in implementing the modernized EMIS to the education sector by user:

→ **Department of Basic Education**
Monitoring and planning for basic education needs in South Africa.

→ **Provincial Education Departments**
Monitoring and planning for basic education needs within the province.

→ **District managers**
Improving educational access and retention, providing management and professional support to schools.

→ **Intergovernmental departments and regulatory bodies**
Providing input into planning and monitoring in other government departments (e.g. the Department of Home Affairs, South African Social Security Agency, South African Council for Educators).

→ **School governing body**
Governing and monitoring school incidents, strategizing to meet set goals and targets.

→ **School principals and heads of departments**
Making informed decisions, monitoring school performance and curriculum coverage, reporting and managing daily running of school.

→ **School administrator**
Performing school administrative tasks and activities to effectively and efficiently manage the school.

→ **Educators**
Performing educator administrative tasks (curriculum progress, learner result management, learner attendance, etc.).

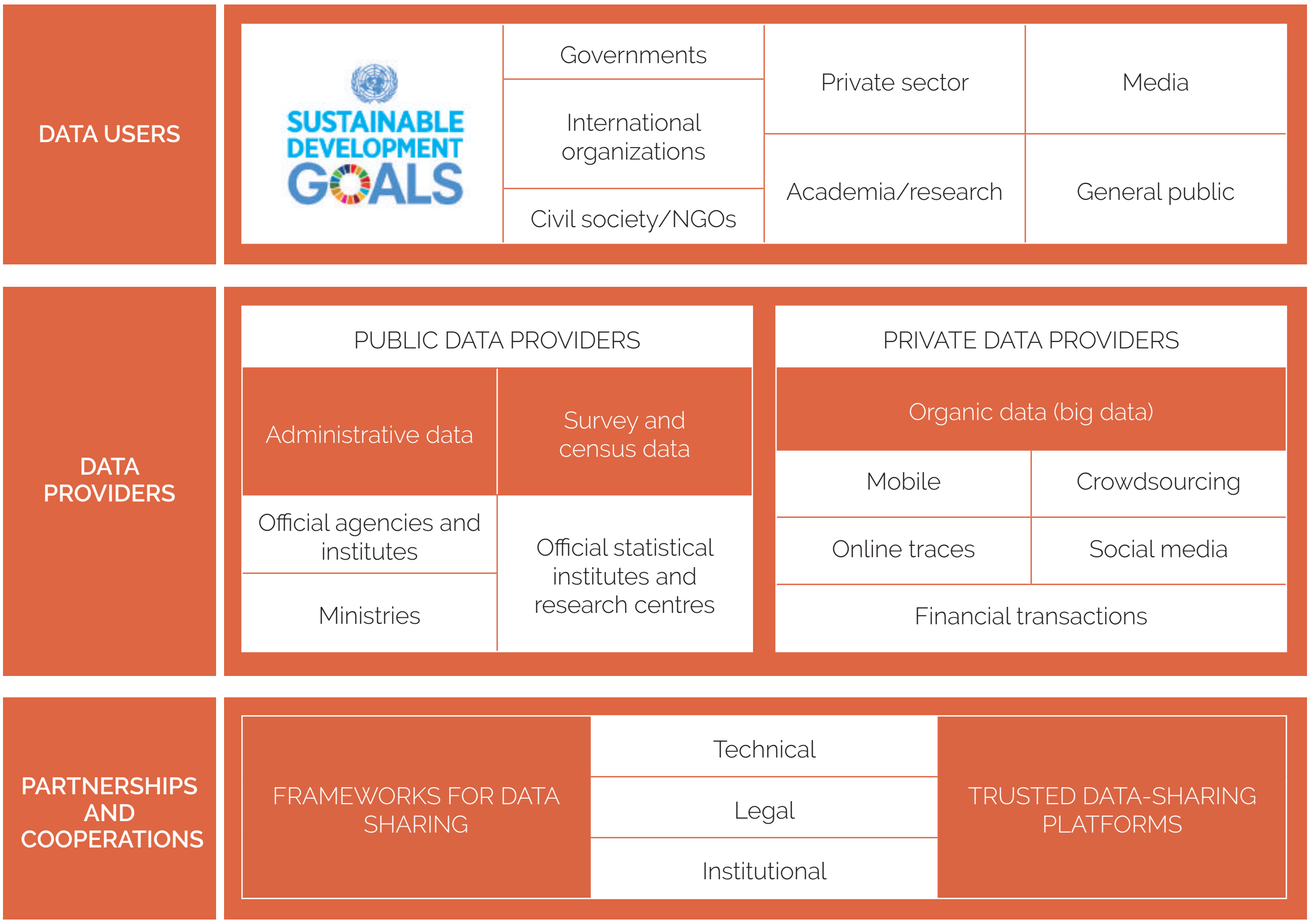
The SA-SAMS evaluation also estimated the amount of time and money saved in South Africa by investing in the open data platform. By moving change management processes to the cloud, an estimated US$13.4 million and 500,000 hours were saved. The administrative burden was reduced by an estimated 10 million hours and US$160 million. Ensuring compliance to South Africa's data privacy protection act by securing data through encryption from device to database prevented costly data ransom threats. By eliminating duplicate learners through an online centralized learner database, 10,500 hours and US$267,000 were saved. Introducing a digital capturing tool for daily learner attendance and lesson registration saved approximately US$160 million and 45 million hours. Finally, configuring South African education systems to align with international reporting standards saved an estimated 240 hours.

Beyond the Department of Basic Education, South Africa's Fourth Industrial Revolution centres, which are designed to train master trainers to develop peers, are piloting twenty-first-century skills, including cybersecurity, in an effort to modernize the TVET system.

Promising practice: Building an ecosystem of educational data in Brazil

The Brazilian data ecosystem of educational statistics has been developed through interinstitutional cooperation among government agencies, research institutions, and civil society organizations including non-profit and private sector organizations (Miao *et al.,* 2022; ITU, 2023). Among the official statistical data producers, the national statistical office (the Brazilian Geography and Statistics Institute [IBGE]) plays the role of coordination of the national statistics.

Figure 7. Statistical data production ecosystem



Source: WGD4L experts from the Brazilian Network Information Centre (NIC.br), 2023.

In the field of education, the National Institute of Educational Studies and Research Anisio Teixeira (INEP) is a federal organization linked to the Ministry of Education, with the mission to coordinate the application of large-scale national educational assessments and to provide evidence for the formulation of educational policies at different levels of government. INEP is also responsible for the Census of Basic Education, which collects a limited number of indicators on the presence of digital technologies among schools.

Regarding information and communication technology (ICT) statistics, most of the national representative research carried out in Brazil comes from the Regional Center for Studies on the Development of the Information Society (Cetic.br), a department of the Brazilian Network Information Centre (NIC.br) that also implements the decisions of the Brazilian Internet Steering Committee (CGI.br). The responsibilities of NIC.br also include the production and dissemination of reliable and representative statistical data on access to and use of digital technologies in the various segments of society, including education. The work carried out by Cetic.br has gained national and international prominence due to the quality and innovative methods used for producing statistical data on ICT. All research projects conducted by Cetic.br are fully funded by NIC.br through the financial resources from the registration activities for domain names under the .br country code top-level domain.

The research process at Cetic.br is structured in a participatory and multi-stakeholder approach (Figure 7). The main instrument for measuring the adoption of digital technologies among principals, director of studies, teachers and students is the ICT in Education survey, conducted annually since 2010 (Brazilian Internet Steering Committee, 2022). Educators and representatives from civil society, academia, government, private institutions, and international organizations participate in both the planning stages of the research and the analysis and dissemination of the indicators to society. The ICT Education survey indicators are widely disseminated, so that mainly education policy-makers, educators, students and their families can use the data and analyses generated based on them.

More recently, another interinstitutional arrangement is monitoring connectivity in Brazilian public schools. The Connected Education Internet Measurement System is aimed at mapping Internet quality and connectivity in Brazilian public schools (Broadband Commission, 2020). The system consists of a software package to measure the quality of Internet connection in schools (download and upload quality, jitter, latency and packet loss) and a data visualization portal[5] that allows public managers and educators to monitor the offer of Internet connection in the group of schools of an education system, and in each school individually. This system was fully developed, implemented and funded by NIC.br through its Center for Studies and Research in Network Technology and Operations (Ceptro.br), in partnership with Cetic.br. The system, including all software components, tools and a portal were developed and made available in kind to schools, educational departments and to the Brazilian Ministry of Education.

The NIC.br Internet connection quality measurement project is the basis for the work carried out by civil society institutions in conjunction with public decision-makers in order to promote the expansion of connectivity in schools, such as the Interinstitutional Group on Connectivity in Education (GICE), formed by government agencies, operators, regional providers, technology companies, associations and third sector organizations (GICE, 2021).

Cybersecurity in education information system architectures

Cyberattacks are dramatically on the rise, with the education sector experiencing a 44 per cent increase in attacks between 2021 and 2022 (Check Point Research, 2022). These attacks come with a cost. According to IBM's 2023 annual report, the global average cost of an education sector data breach was measured at a staggering **US$3.65 million** per critical infrastructure breach. Despite these figures, education organizations have lower rates of cyberinsurance coverage against ransomware when compared with the global average (Hess, 2022; Lauver, 2022).

The education sector also has not invested heavily in training the education workforce on cybersecurity. Only half of countries have standards for developing general teacher ICT skills, and far fewer have teacher training programmes that specifically cover cybersecurity (UNESCO, 2023*b*). Given that the education sector accounts for 5 per cent of all

5 Diagnóstico da Conectividade na Educação is available at https://conectividadenaeducacao.nic.br/.

ransomware attacks and 30 per cent of security breaches, this asymmetry of high rates of attack and low levels of cybersecurity skills in schools indicates a lack of preparedness to prevent against or recover from future attacks on education institutions (UNESCO, 2023*b*).

> **Cyberattacks are on the rise in the education sector, which accounted for 5 per cent of all ransomware attacks and over 30 per cent of security breaches in 2022.**
> (UNESCO, 2023*b*)

With their complex network of staff, students, devices and digital platforms, the size and openness of education systems make them a prime target for cybercrime. Common cyberthreats include ransomware, phishing attempts, IoT attacks, outdated software and data breaches. Preventing and recovering from such cyberattacks puts a price tag on the safe, secure, and effective use of digital technology in education. However, if education systems are to uphold all students' rights to privacy, security and indeed to education, then ensuring cybersecurity in learning spaces is a price that must be paid. The sector must take swift action to adapt to the growing need for digital protection (Mahendru, 2022).

As dependency on digital technology grows in education, so too should the measures to ensure its safe, secure and uncompromised use. It is essential to have a multi-stakeholder, multilateral approach that enables cooperation between the public policy, education, technology, and legal sectors. This approach would ensure that data security is a key consideration in the design and implementation of teacher training programmes and digital learning programmes so that learners in the global South can benefit from digital tools without sacrificing their rights to privacy, safety, and indeed, to education.

When a school's digital ecosystem is compromised, blended and online learning can be interrupted, administrative functions prevented, and the privacy of sensitive student information compromised. When educational provision is online, such as during crisis-triggered school closures or in online courses, learning can come to a crashing, costly, and complicated halt. Moreover, in many schools and universities, students and staff often access educational networks with personal devices, which may not be embedded with the level of built-in protections necessary to prevent attacks

> **Cyberattacks strain education budgets, with the average critical infrastructure data breach costing US$3.65 million. Ransom fees to recover stolen data are often only partially successful, with one study of over 500 education decision-makers from 30 countries recovering just 68 per cent of data after paying the fees.**
> (IBM, 2023)

IBM reports that nearly half (45 per cent) of data breaches globally occur in the cloud. Given the growing use of cloud services for teaching, learning and administrative purposes, this is a worrying figure. When data is transferred within the cloud using an API, "a hacker who breaches such an API can hijack any apps that use the interface to collect data", which could include an administrative software, such as a LMS or another learning platform. Moreover, personal data stored on the cloud can be in numerous jurisdictions, which results in low levels of prosecution of cybercrime internationally, according to a UNESCO report (UNESCO, 2023*b*).

Table 4. Promising practices of cybersecurity strategies in the education sector

Australia: Implementation of university cybersecurity strategies

The higher education sector in Australia is working in partnership with the Government to mitigate risks and promote Australia as an attractive international research and education partner. The Australian Government and the higher education sector jointly formed the University Foreign Interference Taskforce (UFIT) in August 2019 to increase awareness of, and resilience to, the risk of foreign interference in the Australian university sector. The guidelines released in 2019 are globally recognized as a world-leading innovation to help protect Australian research and education assets in an increasingly complex environment. The overarching principles applied when developing these guidelines are:

- Security must safeguard academic freedom, values, and research collaboration.
- Research, collaboration and education activities remain mindful of the national interest.
- Security is a collective responsibility with individual accountability.
- Security should be proportionate to organizational risk.
- The safety of our university community is paramount.

These guidelines are intended to be applied proportionately to the risk at each institution and are not intended to introduce unnecessary burdens on universities. Universities are encouraged to consider the guidelines, identify their own highest risks to help prioritize resourcing, and apply mitigations that are appropriate to their specific risks.

United States: CTE CyberNet and NICE

The **CTE CyberNet programme** aims to prepare high school students with the range of knowledge, skills, and abilities to enter cybersecurity career and educational pathways by supporting the development of rigorous Career Technical Education (CTE) programmes and education of the teaching workforce. The CTE CyberNet is a network of teacher professional development intensive academies led by two-year and four-year post-secondary institutions designated by the National Security Agency (NSA) as Centers of Academic Excellence (CAE).

The CTE CyberNet was strategically designed as an education ecosystem development initiative to both develop the cybersecurity skills of high school teachers, and create a community support structure or local "ecosystem" to increase persistence and sustainability of the academies. The objective of the CTE CyberNet is to rapidly increase the capacity of high school teachers to teach CTE cybersecurity courses, increasing the number of teachers certified to teach cybersecurity, increasing access to CTE cybersecurity programmes of study for students in underserved communities, and expanding the pathway to cybersecurity programmes at technical and community colleges.

The **National Initiative for Cybersecurity Education** (NICE) is a partnership among the Government, academia, and the private sector focused on education, training, and workforce development that will strengthen the cybersecurity posture of organizations. The NICE Workforce Framework for Cybersecurity (NICE Framework) (National Initiative for Cybersecurity Careers and Studies, 2022) provides a set of building blocks for describing the tasks, knowledge and skills that are needed to perform cybersecurity work performed by individuals and teams. Through these building blocks, the NICE Framework enables organizations to develop their workforces to perform cybersecurity work, and it helps learners to explore cybersecurity work and to engage in appropriate learning activities to develop their knowledge and skills. NICE is led by the National Institute of Standards and Technology (NIST) in the United States Department of Commerce.

Costa Rica: Safeguarding education data with privacy and security

Costa Rica's Ministry of Public Education has implemented the Sistema de Información y Gestión Educativa (SIGED) to securely collect and manage student data. SIGED focuses on privacy, consent and data security to provide valuable insights for educational planning, policy development and resource allocation (Arias Ortiz *et al.*, 2021). The system incorporates strict protocols, encryption mechanisms, and legal frameworks to protect student information and mitigate the risk of unauthorized access or breaches. To uphold privacy and consent, SIGED adheres to Costa Rica's Ley General de Protección de Datos Personales (General Law on Personal Data Protection) (OneTrust Data Guidance, 2022). Consent is obtained from parents or legal guardians at the time of enrolment and can be updated or withdrawn as per individual preferences.

SIGED manages a substantial volume of educational data, including academic performance, attendance records, demographics and social indicators, for approximately 1.2 million students enrolled in public schools. Data security measures such as encryption, access controls and regular system audits are implemented to maintain confidentiality and privacy. Data sharing among educational institutions, policy-makers and stakeholders is facilitated by SIGED, informing evidence-based decision-making and educational planning. The Ministry of Public Education ensures compliance with privacy regulations and data protection standards through regular monitoring and assessments (OECD, 2021*b*).

SIGED's implementation in Costa Rica highlights the significance of privacy, consent, safety and security in education data governance. By collecting, managing and protecting student data in alignment with legal frameworks and best practices, SIGED enables data-*driven* decision-making while safeguarding student information.

England: CyberFirst

England (United Kingdom) offers a model for ensuring data privacy and security in the schooling system. CyberFirst is a government cybersecurity learning programme that supports the development of an advanced digital economy for secondary school children and beyond. Since 2016, England has seen 55,000 students engaged with the courses and the CyberFirst Girls programmes, and 1,100 successful CyberFirst Bursary applicants. Its Levelling Up Education standards programme has a goal of improving cybercapability and cybersecurity in schools and colleges to deliver high-quality remote education.

Conclusions: Cooperation and partnership between government and industry

Multisectoral partnerships play a central role in the success, security and sustainability of EMIS and other education data systems. As data systems become increasingly complex, comprehensive and integrated, the cost of storing, processing and managing increasingly large amounts of data carries a heavy financial and human resource need. Strong relationships with private sector providers for both technical expertise and cost-effective and national-scale partnerships can greatly enhance the transformative potential of education data systems. However, for collaboration to be safe, effective and mutually beneficial, there must be a clear and comprehensive legal framework, incorporating well-defined policies and regulations that support EMIS implementation and regulate the involvement of third-party providers.

→ **Financing**
To build strategic and long-lasting relationships between public and private sector stakeholders for information sharing and coordination, education data systems must be primarily grounded in national financial resources allocated for education data systems, with dedicated budgets and robust funding mechanisms.

→ **Integration**
Governments should implement linked unique IDs for students and teachers to ensure accurate tracking and management of education data. Governments should provide "single sign-on" authenticated solutions that allow for seamless integration between industry solutions and government-provided solutions, simplifying access for administrators, educators, and students and their families. Industry stakeholders should ensure that their solutions can be seamlessly integrated into the digital learning ecosystem, allowing for efficient data exchange and interoperability.

→ **Interoperability**
Governments stakeholders should publish interoperability standards that enable the smooth flow of data from LMS to central student information systems, ensuring efficient and secure data management and utilization. Governments should steer collaboration with industry stakeholders to develop interoperability standards, including data standards and technical standards, to facilitate smooth integration and data sharing.

→ **Innovation**
Industry stakeholders should showcase impactful use cases, backed by rigorous evidence from empirical studies, where data has significantly improved the quality and effectiveness of education, and actively publicize these success stories.

→ **Accessibility**
Industry stakeholders should make their data accessible to researchers and third parties under appropriate privacy-respecting statistical conditions, or work with governments to ensure such accessibility, facilitating further analysis and insights. Industry stakeholders should develop secure, age-appropriate and user-friendly solutions that enhance the understandability and actionability of EMIS data for administrators, educators, and students and their families, supporting effective use and interpretation of the data.

→ **Cybersecurity**
Governments should work together to enshrine the right to personal data protection as a fundamental human right. International efforts should be directed towards improving the coordination of reporting cyberrisks and establishing a regular data collection mechanism on cyberattacks against education. Meaningful multi-stakeholder partnerships are required to ensure that learners have opportunities to develop the cybersecurity skills and knowledge demanded by modern workplaces. This can involve linking education programmes with industry-led training programmes and internships, as well as providing support for entrepreneurship and innovation initiatives in the domain of cybersecurity.

→ **Agility**
Governments and industry should work together to provide support to education system stakeholders when problems or challenges arise, assisting in resolving issues related to data integration and utilization. This includes the swift mitigation of risks and cyberthreats.

→ **Actionability**
Governments should gather a variety of useful information, including learning achievement data, and use this information to provide immediate feedback to stakeholders, highlighting the value of the data and its impact on education outcomes. Industry stakeholders should provide education systems and stakeholders with actionable information and data derived from their solutions, enabling evidence-based decision-making and targeted interventions.

Critically, using data to improve learning, teaching and management can only be realized if education stakeholders have the skills and competencies to interact with these data systems in safe, purposeful and meaningful ways. The next section will explore the various data literacies needed to untap the transformative potential of data for learning in an age where data systems are growing in technical capabilities, analytic power, user friendliness and industry control.

3

Data skills and competencies in education

Defining data literacy

Like digital literacy, the concept of data literacy is fluid, changing form as flows of data change in power and application. Broadly, data literacy is a multilayered concept comprising a combination of technical-statistical and analytic-narrative skills for ethical and effective collaboration between people and data-*driven* tools. It is defined by UIS as "an individual's ability to access, manage, understand, integrate, communicate, evaluate and create information safely and appropriately through digital technologies" (Antoninis, 2018). However, this definition deserves revisiting in today's world where data-fuelled tools are growing in power and prominence. It is critically important for all people to be aware of how algorithms work, how they are used in digital services, and how they impact our communities and social relationships.

In 2022, UNESCO conducted a mapping of government-endorsed kindergarten to grade 12 AI curricula, which further developed the concept of data literacy and its component knowledge and skills, to arrive at an operational definition of AI literacy. AI literacy combines the information literacy of the definition above with algorithm literacy, or "the ability to understand how AI algorithms find patterns and connections in the data, which might be used for human-machine interactions" (UNESCO, 2022*a*). A rigorous definition that clearly outlines the competencies required to be data and algorithm literate is planned for release by UNESCO later in 2023, with the aim of using this definition to develop frameworks of AI competencies for teachers and students and to guide the empowerment of individuals in navigating around an AI-rich environment, making informed decisions, and actively shaping its responsible and beneficial use.

> Globally, only 11 out of 15 governments surveyed have implemented curricula that develop AI and data competencies.
> *(UNESCO, 2022a)*

Education and training systems play a leading role in developing the knowledge and competencies that constitute being data literate in an age of algorithmic decision-making. Despite its relevance around the world, data literacy is not commonly integrated into general teacher training curricula. However, a clear asymmetry of skills exists as, to date, there are only 14 AI curricula which have been developed and implemented by 11 out of 51 governments surveyed. (UNESCO, 2022*a*). Even where elements of data literacy development are present in national training programmes, these competencies are typically relegated to upper levels of science, mathematics or statistics classrooms. For every student to understand and navigate the increasingly datafied world as a critical evaluator and co-pilot of any algorithmic decision-making, teachers need to be empowered to teach data literacies across subject matter and grade levels.

Data competencies for teachers and lifelong learners

For teachers to be able to properly teach data literacy, they need basic digital literacies to understand how to navigate the Internet and how they can create, transfer and store digital data. However, only 54 per cent of countries have digital skill standards, which are often defined by non-state actors rather than national governments (UNESCO, 2023*b*). ITU breaks down digital skills into five categories: communication/collaboration, problem solving, safety, content creation and information/data literacy. Overall, ITU (2021) estimates that the average proportion of individuals in developed countries with basic digital skills is 65 per cent, with 46 per cent in developing countries. When considering standard intermediate skills , the numbers fall significantly to 49 per cent in developed countries and just 20 per cent in developing countries. Focusing on the information/data literacy category, in half of all countries, less than 28 per cent of individuals have these skills. In one in every four countries, the percentage decreases to below 17 per cent of individuals. With these low levels of individual skills or standard-setting instruments, it is clear that there is much work to be done to empower teachers to ethically and effectively use education data worldwide.

> **Only 65 per cent of people in developed countries – and 46 per cent in developing countries – have basic digital skills (ITU, 2021). In half of all countries, less than 28 per cent of individuals have data literacy skills.**

Many digital skills development frameworks are based upon a belief in learning by doing, or that by using technology, especially with the support of a skilled instructor, one will acquire digital skills. The same logic could be applied to data literacy. If the learner participates in the production and analysis of personal data, and treats their own data as a way of critically exploring their beliefs, values, responses and social identities, then they could improve their data literacy and move "beyond the typical schooling practices of restating and critique" (Sheridan and Rowsell, 2010, p. 111). The learning-by-doing approach puts a heavy burden on teachers to teach data literacy by implementing best practices, which requires them to possess the necessary digital and data literacies themselves – skills which the majority of teachers globally do not have.

As a result, the training teachers receive to engage with student data in informed and productive ways is of vital importance. Frequently, school districts or systems will purchase a one-off, vendor-based training with a technology purchase that may include LMS, assessment and early warning systems, dashboards, and other applications depending on the system's educational objectives (Mandinach and Gummer, 2021). Often these trainings focus on how to access data, rather than on how to *use* the data to inform pedagogy. If teachers are not supported by their schools and wider development programmes and training systems to understand how to contextualize the information their students produce to inform their teaching practice, then learning data risks being misused, leading to simplifications or misinterpretations.

In education systems at all ranges of digital and data system development, teachers interact with many sources of learner and learning data, some of which they may collect and others they may not. For example, they may collect data during teaching and learning processes, and assessments, as well as the data drawn from learner interactions with digital environments. This learning data can be collected by hand, recording the achievements of learners in typical pen-and-paper activities, but might also be collected automatically in digital systems that include monitoring functionalities, and may be underpinned by powerful data analytics and AI. For example, teachers may analyse a student's digital learning report to identify a particular skill the student may need additional support developing. They may use the data generated by digital technologies to reflect on the best teaching strategies to employ for specific lessons. By looking at trends in class performance, they could differentiate their instruction and scaffold their lesson plans accordingly. They may use data-*driven* tools to lower the burden of administrative tasks such as monitoring attendance, tracking homework assignments, or entering assessments into gradebooks.

Given this wide range of uses, it is essential that teachers have a diverse set of competencies – technical, legal, social and ethical – to support their purposeful interactions with education data.[6]

→ **Ethical data-*informed* instruction**
Teachers should be aware of the various forms of personal data used in education and training. Teachers should be able to collect, analyse and interpret student data to inform their instructional practices. Teachers should be able to interpret the data and evidence available in order to better understand individual learners' needs for support. This includes through classroom-based assessments, manual or tech-based tracking of student progress, identifying learning gaps, and adjusting teaching strategies accordingly to meet individual student needs. Across use cases, teachers should consider potential analytic biases caused by various data-related factors, including algorithms, censorship and limitations in their own data literacies.

6 The teacher and lifelong learner competencies featured in this report were predominantly sourced from Vuorikari *et al.* (2022) and its subsequent analysis by Kivinen *et al.* (2022).

→ **Assessing data quality**
Teachers need to understand how to assess the quality and reliability of the data they use (through assessments and engagement in digital environments) as part of their teaching. They should be able to critically evaluate education data, educational tools and research, and other sources to ensure data validity and make informed decisions, not only about instruction, but also about privacy, age-appropriateness and legal compliance. Teachers need to critically assess and discuss the value and validity of different data sources, as well as the appropriateness of established methods for data analysis.

→ **Awareness of data governance architecture and legal compliance**
Teachers also need awareness around the main principles of data governance and ownership to inform and protect their learners' privacy and rights. Teachers should be able to identify both the positive and negative implications of the use of data, and weigh the benefits and risks before allowing third parties to process personal data. They should know the legal and policy frameworks applicable to their school contexts, and how this implicates their rights to informed consent, who has access to student data, with whom it is safe to share data, how access is monitored, how long data are retained, and how data can be deleted.

→ **Understanding the impact of data use on humans and human rights**
Teachers should know how a given digital system addresses the different social objectives of education and how data profiling can influence societal decision-making. Teachers should be able to consider the impact of data use on the student community. Teachers should be able to explain how a data use can benefit all students, independent of their cognitive, cultural, economic or physical differences. Teachers should be able to consider the impact of data use (learning analytics, automated feedback and assessment, and progress monitoring) on the development of student self-efficiency, self-image, mindset, and cognitive and affective self-regulation skills.

→ **Facilitating students' data literacy**
Students need skills to analyse and interpret data relevant to their learning. This includes understanding statistical graphs to understand their learning trajectory and arguments for pursuing particular learning pathways. Students need to be aware that digital systems collect and process multiple types of their data (e.g. personal data, behavioural data and contextual data) to create user profiles. They also should support children to know how their data are then used, for example, to suggest learning paths or to predict their success (or failure) based on algorithms. Teachers should give students practical and experience-based advice on how to safeguard their personal data, and mitigate risks related to safety and privacy in digital environments. Teachers should be able to use different activities and projects to help students learn about the ethics of data use in education, including how data systems can direct learning and impact human rights.

→ **Data-*informed* collaboration**
Data literacy also enables teachers to collaborate effectively with colleagues by sharing best practices, analysing data collectively, comparing and critically evaluating the credibility and reliability of data from digital environments, and collaborating on interventions to improve learning experiences and foster student achievements. Teachers should be able to present data in meaningful and accessible formats to facilitate student, parent and leadership understanding of classroom activities.

Box 7. Eco Ambassador Summer Program

The Eco Ambassador Summer Program, led by Columbia University's Climate School and SDGs Today of the Sustainable Development Solutions Network (SDSN), provides 10 weeks of training in data and technical skills (Geographic Information System [GIS], AI, etc.), the SDGs and digital storytelling annually. A free and virtual programme, it has attracted students from over 60 countries who use these skills to collect, analyse and use data for sustainability challenges in their communities. The programme inverts the use of data from being dictated by teachers to students towards the student eco ambassadors who use the data to narrate their own climate stories, which can then be discussed or facilitated in classrooms.

From a lifelong learning perspective, all learners need skills to analyse and interpret data relevant to their learning. This includes understanding statistical graphs to understand their learning trajectory and arguments for pursuing particular learning pathways. They need to understand how learning analytics, automated feedback and assessment, and progress monitoring, may direct their learning experiences, and how their data is used in their digital learning environments and beyond. Beyond the analytic dimension, learners – like teachers – need to also possess a broad range of competencies to minimize the risks of data use in learning settings, empower learner agency over their data and digital learning, and spur sustainable innovation through data-powered tools.

→ **Managing digital learning profiles**
They should be aware of the benefits and dangers of integrated data systems that can track learning data over their lifetimes, understanding the opportunities this could present for recognizing competencies and learning accomplishments, as well as the repercussions it could have on their self-perception or algorithmic limitation of their future learning or employment choices. They should know the privacy policies of the LMS they engage with, so that they understand how their data is used to influence their learning experiences and by whom.

→ **Protecting personal data**
In the digital age, students in particular need to be aware of the ethical and legal considerations surrounding the collection and sharing of their personal data as a result of their interaction with digital learning environments, including LMS, EdTech, phone-based learning apps and participation in large-scale learning assessments. In short, anywhere where personal information is gathered and could be used and stored by a third party, students should be aware of the legal protections applicable in their contexts that safeguard their human rights. This involves an awareness of compliance procedures related to data privacy, informed consent, transparent data use and responsible cybersecurity practices in their school environments. They should understand how to use and share personally identifiable data and information while being able to protect themselves and others from the potential risks of personal information being stored by third-party users.

→ **Data-*driven* problem solving**
As part of project-based learning, students also need data literacy skills to solve real-world problems. They should be able to collect and analyse data, draw insights, and propose solutions based on evidence. This fosters critical thinking, creativity and an understanding of the potential value of data in decision-making.

→ **Environmental impact of data use**
Students, like teachers, and indeed all people, should understand the environmental impact of everyday digital practices that rely on data transfer, which produces carbon emissions from devices, data centres and network infrastructures. In particular, they should understand the energy-intensive processes that power the digital learning environments which employ AI.

Beyond teachers and students, there are additional school-level and community-level responsibilities needed to responsibly unlock the potential of purposeful data use to improve the educational experiences. Across all cases at the school level, skills and competencies for data governance directly include the active protection of the data rights of learners, teachers and families, including informed consent, protecting access to data and opting out of specific data practices. While there may be an appointed IT person charged with technical troubleshooting, IT staff support and cybersecurity for the school community, it is important to note that this may not be feasible in low-resource contexts where governments are already struggling with school maintenance costs and paying teacher salaries. Specific requirements and skillsets at the school level would also depend on the size and complexity of the school's IT infrastructure and the resources available to recruit a dedicated IT officer.

School uses of education data are part of a larger governance architectures at regional and national levels of education and training systems. At the system governance level, decision-makers require additional data competencies to support the protection of the rights of learners and the education workforce, ensure the integrity and usability of education data used to support decision-making, and improve education system efficiency, which is needed to expand access to quality educational opportunities, particularly for underserved population groups.

Data competencies for education system governance

There is no universally standardized definition of data governance specifically tailored to the education sector. However, data governance generally refers to the establishment and enforcement of policies, procedures, principles, standards and practices to effectively and responsibly manage and monitor data use within educational institutions and across diverse education stakeholders.[7] It involves the development of frameworks and strategies to maximize the value, quality, security, privacy, ethics and accessibility of data, while ensuring accountability for compliance with relevant laws, regulations and ethical considerations. The definition also takes into account notions related to data subject rights; protecting the privacy of individuals; data security and incident response; data protection by design and default; lawfulness, fairness and transparency; accountability and data standards; benchmarking and auditing; and data protection impact assessments (DPIAs).

Education authorities (and individuals) can leverage the latest data management, security and governance practices, and ensure efficient and responsible data management within the education sector that is oriented towards protecting and benefiting learners. Within any typical information management system, capacities for data collection, accuracy, analysis and interpretation at a high level are needed. However, the human capacity to engage with high-quality digital data, delivered in real time through relevant technological platforms, is also essential for system functionality and effective decision-making. As a result, today's education leaders should possess a combination of technical knowledge, legal understanding and interpersonal skills. Some of these include:

→ **Knowledge of relevant laws, regulations and standards in the education sector**
First and foremost, education data governance should ensure that there are clear sector-specific standards for the safe and transparent use of all learner-level data, that can be explained to all education stakeholders, from learners to EdTech providers. Likewise, education data governance should regularly ensure that evolving technologies used in education systems meet regulatory requirements, promote data security, privacy, transparency and fairness in data use, and mitigate all risks related to data misuse. As a result, education leaders need knowledge of data protection regulations (e.g. the GDPR), industry frameworks (e.g. ISO 27001) and any specific compliance requirements for educational institutions. They should demonstrate a commitment to staying updated with the evolving landscape of data privacy and protection laws, regulations and best practices through ongoing professional development and networking.

7 See this report's Annex for the full list of data governance references, and the accompanying job board for data governance in the education sector.

→ **Data system architecture expertise and capacity to analyse operations, identify risks and opportunities**
Data governance should ensure that all technologies interacting with learner-level data, and in particular those provided by industry vendors, are optimized to generate an education data architecture supporting safe and compliant data collection, analysis and integration across different education planning, management, reporting and delivery functions. The resulting data outputs can support decision-makers to develop evidence-based strategies for improving learning outcomes, invest in targeted actions to serve marginalized learners, and strengthening system management efficiency and functioning. The education governance architecture should include individuals with strong skills in data manipulation, data cleansing and data integration using tools such as SQL, Python, R or other programming languages commonly used in data analytics, and who are familiar with data analytics platforms and tools for data visualization and reporting like Tableau, Power BI or Excel. This would enable a deep understanding of education databases, data storage, mining, retrieval, integration and governance practices.

→ **Database proficiency, data pattern analysis and interpretation competences**
Usability of education data should be maintained through all governance levels, with an efficient data architecture that allows for collaboration across functional areas to counter data silos and promote interoperable data management. Such interoperable systems must be regularly audited and assessed for compliance with government standards that protect the user, while providing a streamlined and standardized experience to improve data usability for all learners and the education workforce. Data governance should ensure the quality, reliability, accuracy and trustworthiness of all data used in education decision-making. Such guardrails are vital in the context of increasingly complex and interoperable data management platforms, and in the face of new, powerful technologies which can treat large volumes of data from different sources, such as those powered by generative AI.

→ **Privacy by design principles, risk management, response to security incidents, and vendor management skills**
A strong understanding of information security principles, practices and technologies, including knowledge of the technical aspects of network security, encryption, vulnerability management, access controls, incident response and regulatory compliance. Education leaders should have a knowledge of privacy by design principles and the ability to embed privacy safeguards into systems, processes and applications right from the design phase, which includes assessing the security posture of vendors, establishing security requirements in contracts, and conducting regular audits or assessments to ensure compliance.

→ **Communicating data insights, continuous learning, and team cooperation**
Education decision-makers need strong communication skills to effectively communicate data insights to various local stakeholders, including teachers, parents, governing bodies and regional education authorities and school inspectorates. Education leaders should be able to present data in meaningful and accessible formats to facilitate understanding and support decision-making. This requires knowledge of the education sector, including understanding of educational data, performance metrics and the factors that impact educational outcomes. Decision-makers in the education sector should keep up with the latest trends, tools and methodologies in data analytics, remaining adaptable to new technologies and evolving data practices in the education sector. This competency includes a willingness to learn and explore new techniques to improve data analysis and insight generation.

Over the past few decades, the increasing diversity of data sources from stakeholders within the education community has not always been reflected in top-level decision-making. Government actors are not routinely leveraging multiple sources for their strategic planning and decision-making processes. This is partly due to the education system's low absorption capacity for integrating and applying the data that has been produced, and to weak human resource capacities for adjusting management and operational processes to improve service delivery.

Box 8. CNIL: Developing the ethics of digital education

Commission Nationale de l'Informatique et des Libertés (CNIL) is the French data protection authority and is an independent government agency. It is responsible for ensuring the protection of public and private personal data contained in paper and computer files and processed during operations. This responsibility is actioned across four missions such as information and rights protection, compliance support and guidance, anticipation and innovation, and investigations and sanctions. More recently, CNIL has also begun to explore the effects the rapid evolution of new technological tools is having on the organization of life in society, and with regards to data sovereignty as a risk for the reuse of learning data. This has included exploring recommendations around the collective challenge of developing the ethics of digital education.

The agency raises the importance of considering the digital divide and that digital education can generate and reinforce it and that digital technologies should not lock children into systems of predetermination. However, while offering a lucrative market for big tech corporations, digital education also represents a strategic asset for the State, raising the question of governance strategy. For example, in France, the State finances the use of technological tools in schools, which necessitates the consideration of the legal and ethical limits placed on the sharing and reuse of data by the State, as well as by the technology companies implementing the tools. This is while considering that education data is not like other data, as it mainly concerns minors, and contains significant information about their private life. In Europe, this consideration currently comes in the form of the GDPR. CNIL discussions also highlight the importance of cybersecurity, with education data also exposing children to risks such as identity theft, cyberbullying and the commercial exploitation of data, that raise further online security concerns for learner and parent data.

CNIL posits that training and the development of an informed digital culture mark key aspects of successful frameworks for governing digital education ecosystems. This means educating learners to maintain control of their data by demystifying these new technologies, empowering young people with the strategic skills needed in a digital media society and helping adults understand the economic model they are subject to.

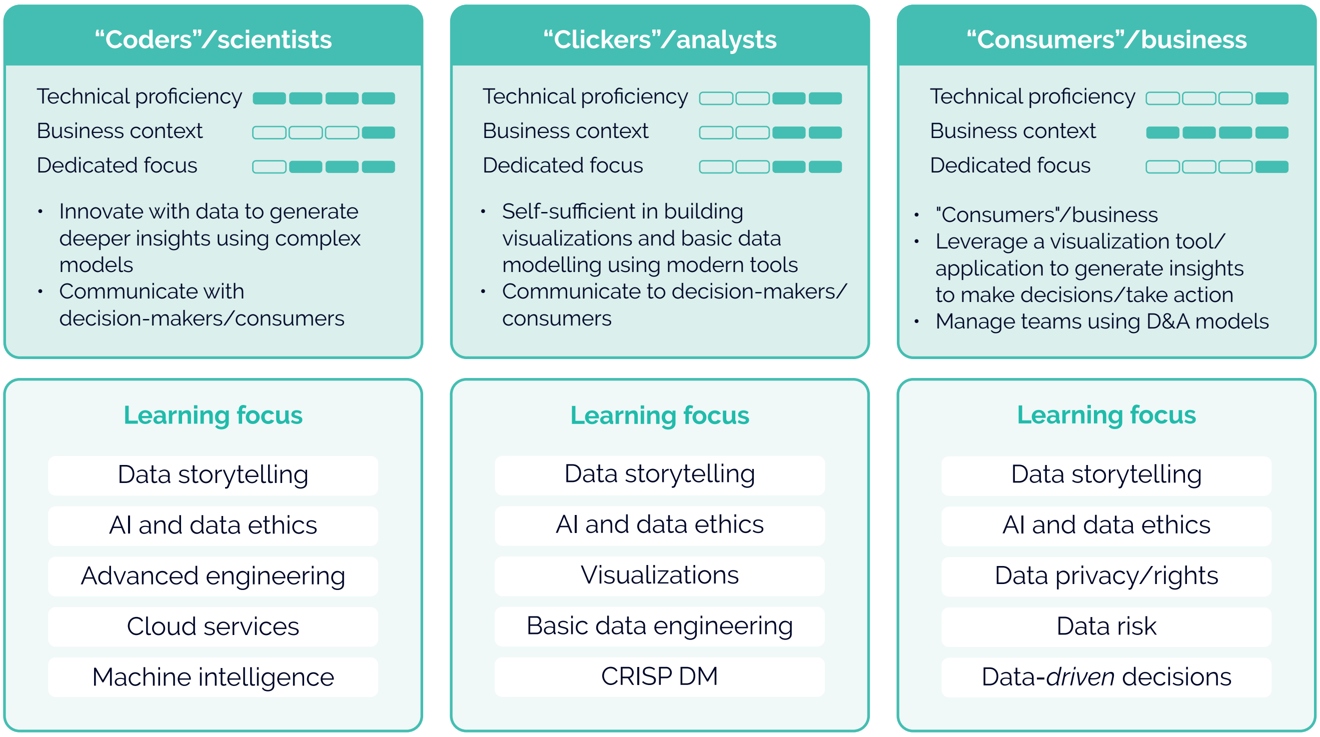
Data competencies for the world of work

From the corporate perspective, many companies have developed their own data skills frameworks to scaffold workforce development and growth through upskilling and reskilling. Many of these frameworks are used by governments, in particular within the education and training sector, in an attempt to strengthen links between the skills cultivated in learning settings and those in high labour-market demand (UNESCO, 2023*b*). However, corporate frameworks for data literacy and programmes to develop such skills are not always openly available for the public sector. Strengthening alliances between industry and education systems could improve the labour-market relevance of the training courses that education systems develop.

For example, KPMG, a multinational professional services network, developed a data literacy programme that helps its own professionals to become comfortable working with data and analytics. To gauge how people progress through this circle, KPMG classifies individuals into three types of learner to determine the required level of data skills:

"coders", "clickers" and "consumers". The common denominator across the three tiers is the standard goal of public engagement for democracy and empowerment through critical understanding of data uses and abuses.

Figure 8. KPMG's data skill needs across three personas



Note: CRISP DM: Cross-industry standard process for data mining.
Source: Presentation to WGD4L by Robert Parr on 5 July 2022.

At the highest level, **coders** are the data scientists on the frontier of generating insights using complex models. Their learning focus is data storytelling, as they tend to have a technical background and may need to develop skills to better communicate complex insights derived from machine learning and deep learning. **Clickers**, or data analysts, are those whose primary job is not to code, but rather to apply data science in business contexts. Their learning focus may be on how to weave data together from different sources credibly and ethically, based on an understanding of basic methodologies for data mining, modelling and visualizing. Finally, **consumers** need training in storytelling, AI and data ethics, and data privacy and rights, for though they may not be proficient data scientists, they require accountability as managers, and must understand the deeper questions to ask of data.

Overall, all private sector actors should have knowledge of relevant laws, regulations, and standards in the education sector is vital. This includes data protection regulations (e.g. the GDPR), industry frameworks (e.g. ISO 27001), and any specific compliance requirements for educational institutions. Industry leaders should be able to articulate complex security concepts in a clear and concise manner, and effectively engage with stakeholders at all levels, including decision-makers and education staff, about security best practices and collaborating with cross-sector departments to implement security initiatives. Moreover, it is essential that education decision-makers from both the public and private sectors establish good working relationships with educational leaders, teachers, administrators, and external and internal partners, including with chief information officers from other sectors of national government.

Additionally, the issue of the accessibility of data skills and literacies for individuals with special needs should be considered in the design of data systems, especially from private providers. Data attached to students with disabilities is often of a sensitive nature, and some countries, such as the United States, have particular legal requirements related to sharing the data of learners with disabilities, as protected by the Individuals with Disabilities Education Act (IDEA) and the Children's Online Privacy Protection Act (COPPA). Although such legal frameworks seek to prevent the abuse and misuse of student data relating to learners with special needs, obtaining genuinely informed consent with the understanding that sharing such data may negatively impact future educational or employment opportunities remains a challenge (Stahl and Karger, 2016).

Organizations, both educational and commercial, should pay particular attention to dimensions of equity in their data literacy frameworks, in order to address the under-representation of women and individuals from the global South in the field of data science. Globally, the gender gap in digital skills is not wide, but it widens significantly with regard to computer programming and specific data literacies or algorithmic skills. For example, in 50 countries, while 6.5 per cent of men could write a computer programme, only 3.2 per cent of women possessed this competency (UNESCO, 2023*b*). However, with the release of powerful new generative AI tools, curricula targeting data literacy and digital skills should consider the potential shifts in future labour-market demands. With increasingly powerful generative AI tools, coders could face the risk of being outsourced or automated, whereas analysts and roles that require critical thinking and communicating data insights may have higher job security due to ongoing demand.

Schools could partner with open-source data initiatives to allow learners to develop their data skills using real-world problems that they care about. Learning without these raw materials is limited, a fact which further underscores the importance of open data as an open educational resource that can place cost-effective learning materials in the hands of educators around the world. If such processes and materials were combined into a structured curriculum, this approach to improving data literacy by focusing on the classroom level could be scaled up for global impact.

For example, initiatives like the **Use (open research) Data in Teaching project** (UDIT) aim to assist educators in higher education in effectively integrating open research data into their instructional methods. This endeavour involves a range of activities, including offering courses that showcase best practices and illustrative instances of learning engagements centred around the reapplication of open data.

Further, the collaboration between the National Library of Spain and Red.es[8] has given rise to the **BNEscolar educational platform**, which harnesses digital resources sourced from the extensive documentary collection of the Hispanic Digital Library.[9] This online platform boasts a search functionality designed to facilitate the discovery of specific resources, alongside workshops, videos, pedagogical sequences, and interactive challenges, such as an escape game. The content available on BNEscolar is tailored for students at the pre-university level, with a special emphasis on the latter stages of primary and secondary education. This exemplifies the potential of open data as a robust pedagogical tool for nurturing learners' engagement and understanding (Government of Spain, 2019).

8 Red.es is the public corporate body affiliated with the Ministry of Industry, Energy and Tourism. See https://www.red.es/en.

9 See https://bnescolar.bne.es/.

Box 9. Understanding the relationship between literacy and transparency: Who bears the burden of informed consent?

The relationship between literacy and transparency is a complex and important one, as highlighted by the Centre for International Governance Innovation (Hong, 2021). According to their findings, transparency can sometimes be perceived as a form of free labour, burdening individuals with misinformation or disinformation while leaving them with limited means to take corrective action. This perspective suggests that there is an excessive shift of responsibility onto the public, expecting them to understand and navigate complex algorithms, read through extensive terms of service agreements, and even fact-check every piece of news they come across. The consequence is that individuals may be unfairly blamed for lacking technological literacy or showing insufficient concern about privacy matters, while the significant efforts made by certain entities to obscure data collection and utilization practices go overlooked.

However, it is essential to take a more balanced approach to address these concerns. One critical aspect involves advocating for the implementation of age-appropriate and internationally aligned consent practices. By doing so, individuals can be empowered to make well-informed decisions about how their data is used and shared. It is further vital to emphasize the demand for transparency and explicability in the development along with the complete lifecycle of technological systems. The need for clear definitions of transparency, and its consistent application throughout the iterative and learning processes of these systems, cannot be understated. Scholars argue that this is crucial for building and maintaining trust between users and technology.

Rather than grouping all practices together without differentiation, a more nuanced and contextual approach would involve exploring various mechanisms to foster transparency in educational settings. Educating users about the potential demands they can make of the systems they interact with would equip them with essential tools to navigate the complexities of technology and data privacy more effectively. Addressing the intricate relationship between literacy and transparency requires a comprehensive and multifaceted strategy. By actively advocating for age-appropriate and internationally aligned consent practices and placing emphasis on ongoing transparency in technology, individuals can be empowered with the knowledge necessary to protect their privacy and make well-informed choices. This approach aligns with the call for a redistribution of power and a move towards a more equitable information landscape.

Conclusion: Data skills for all

One year following the Transforming Education Summit, the Secretary-General of the United Nations released a new vision for turbocharging collective progress towards the SDGs, called UN 2.0.

The world is halfway through its timeline to achieve the SDGs, but half the world remains excluded from progress. At the heart of UN 2.0 is a “quintet of change”: **nurturing cutting-edge expertise in data**, innovation, digital, behavioural science and strategic foresight.

This quintet of change places data and digital capacities at the centre of the wheel driving the transformative change the world needs urgently. This wheel is turning towards the release of the Global Digital Compact in 2024 during the Summit of the Future, which is intended to outline shared principles for an open, free and secure digital future for all.

To advance progress towards the SDGs and support Member States’ strategic decision-making, the UN 2.0 Policy Brief strongly urges the global community to invest in modern skills, the first of which is data expertise.

4

Key considerations for education data governance

In addition to those identified earlier, many of the challenges and risks related to governing data in education, especially big data, remain radically underexplored. As mentioned in Part 2, the global datasphere is projected to reach 175 zettabytes by 2025, representing a fivefold increase from 2018.[10] This vast amount of data holds immense potential for various sectors, including education and governance, but also raises concerns regarding ownership, governance and the ethical use of data, among others. While data collection can empower educators, improve management and enhance learning experiences, it intertwines with economic and political power, risking harm to learners.

This section of the report will, as a starting point, highlight key considerations around education data ownership, ethics, and financial and environmental sustainability that are critical for any data decision-maker to integrate into a data governance strategy at all levels of the education ecosystem.

Ownership and the common good

A growing concentration of data is in the hands of a few large corporations. As of 2021, the top five technology companies (Apple, Amazon, Microsoft, Alphabet/Google and Meta) accounted for a combined market capitalization of over US$8 trillion. These companies not only collect vast amounts of user data, but also have significant control over data-related technologies and infrastructure. This concentration of data and power raises concerns about the potential for abuse, privacy breaches and unequal access to data-*informed* benefits, including personalized education.[11] Balancing profit-driven data accumulation with data as a common good requires addressing the challenges posed by the data-related power held by a few commercial players and countries concentrated in the global North.

While there are risks associated with how EdTech companies create and use data, these should be assessed in the context of institutional and operational realities. With the proper guard rails and sustained cooperation, EdTech can work with governments on different dimensions of delivering digital transformation in education, which could in turn, complement government efforts to ensure that digital learning options are freely made available to all learners as a common good.

The notion of education as a public good emphasizes the state's responsibility for equal access and social justice. However, the influence of non-state actors in education complicates this definition, prompting the adoption of the "common good" concept to refer to community cooperation, as described in Table 5.

> **Despite the desire to make education a global common good, the role of commercial and private interests in education continues to grow.**
> *(Audrey Azoulay, UNESCO, 2023b)*

10 See Reinsel *et al.* (2018).

11 See Statista (2023).

Table 5. Public and common good

Education as a public good
UNESCO, along with many other civil society and United Nations organizations,[12] have used the term 'public good' to reaffirm a humanistic vision of education, justify the need to safeguard public interest, and "reject calls for increased privatization or commercialization in education" (UNESCO and UNESCO and Collective Consultation of NGOs, 2015, p. 5).
Education as a common good
Education as a common good introduces the values of innovation and community cooperation, or as Locatelli (2018, p. 11) states, "it envisages new and innovative education institutions that can improve quality and efficiency thanks to the empowerment and the greater cooperation with the forces present in society." The sociocultural concept centres community justice and well-being over individual socio-economic investment.
Data for learning as a common and public good
Data for learning in this context, thus, should embody the democratically governed values implied by "public good" and the sociocultural, innovative and collaborative values implied by "common good". For examples of how open data may be considered an embodiment of this concept of data for learning as a common and public good, see more about the Open Data movement (UNESCO, n.d.).

Presently, resource-rich global businesses, foundations, publishers, venture capitalists and technology companies drive data expansion in education. Dominance of commercial data use and ownership models restricts users' autonomy over their data and the ability of governments to open learning data to advance research at scale and comparative evidence to support best practices in education. While it is crucial to assess how industry support can benefit resource-scarce education systems without undermining public control, it is also necessary to consider the operational limitations of public institutions to guarantee the anonymity and reliability of open learning data. Consistently ensuring the privacy of learners would need to be carefully safeguarded in order for open data for learning to be a common good used to spur cross-border research and innovation. However, minimal international cooperation on data governance and regulation at present impedes the realization of data for learning as a common good.

The models around monetization of data for learning are extremely sensitive as learners' data hold value for various stakeholders, from EdTech providers to corporations, political parties and advertisers. Learners' data hold value for various stakeholders, from political parties and corporations to EdTech start-ups and advertisers. To determine ownership of intellectual property related to data, reconciling learners' ownership rights with claims from software providers and EdTech companies is necessary.

> **Human Rights Watch found that of the 164 EdTech products recommended for children's learning during the pandemic, 89 per cent could or did follow children in educational settings or outside of school hours. Tracking technologies installed on learning platforms collected and sent data on children to third-party companies.**
> *(Human Rights Watch, 2022)*

Cross-border and multilateral normative instruments can establish ethical principles for public and private institutions, given the role student data may play

12 These include, among others, the Global Campaign for Education, the Right to Education Initiative, the Global Initiative for Economic Social and Cultural Rights, and Education International (Locatelli, 2018, p.2).

in business plans, which contravenes the United Nations Convention on the Rights of the Child right not to be subject to commercial exploitation (see Article 19).

Multilateral cooperation is also necessary to navigate complexities such as intellectual property legislation, licencing and control over search models. As Part 3 explored, EMIS are increasingly centralized with unique identification numbers that streamline access to education and other government services. However, as these databases grow in scope and granularity, so do concerns over unauthorized access and surveillance. To mitigate these concerns, governance frameworks must prioritize strong data protection and cybersecurity measures, including encryption and access controls, to safeguard personally identifiable information (PII) in education data against data breaches and privacy violations. Clear guidelines should determine data ownership and custodianship, and transparency and accountability mechanisms – such as regular audits and public reporting – play a crucial role in addressing ethical concerns and enhancing trust.

Individuals should be able to access, rectify and delete their data, supported by informed consent mechanisms that provide clear information about data usage and protection. At the micro level, current data governance models and low levels of data literacy prevent children or their parents from being able to do this. At the macro level, cross-border cooperation is an essential step towards clarifying legal ownership and ensuring that learners can access their own data even if the platform provider sits within a different national jurisdiction. Governments, educational institutions and stakeholders should collaborate to develop robust data protection policies. These frameworks should define roles, set data protection standards, and establish protocols for data transfer and storage. Clear agreements and collaborations among stakeholders are necessary to ensure the secure exchange of education data and foster equitable and inclusive educational opportunities across borders.

Box 10. Digital Futures Commission: Blueprint for child rights-respecting data governance

Data are collected from children all day long – at home, in the street, during their leisure time and while they learn at school. The data are personal, even sensitive, and can be analysed to reveal intimate details about each child. The Digital Futures Commission quickly discovered that sharing children's data is fraught with risk (Day, 2021*a*), mainly because data governance is weak (Day, 2021*b*). Through a series of sociolegal investigations and interviews with schools, data protection officers and other experts (Turner *et al.*, 2022), the Commission revealed the unfair burden placed on schools (Turner, 2022) to negotiate contracts with opaque and powerful companies (Hooper *et al.,* 2022), and the lack of data protection compliance of some of these companies (Digital Futures Commission, 2022). The Digital Futures Commission has produced an essay collection (Livingstone and Pothong, 2022) exploring how robust data governance could address the problems of education data processing, while new approaches to data stewardship (Zhao, 2022) could open new possibilities for sharing education data in children's best interests and the public interest (Knight and Hannay, 2022). The collection concludes with a blueprint for child rights-respecting data governance (Digital Futures Commission, 2023) and practice, setting out three priorities for the government and the regulator to focus on: (1) Strengthen existing legal frameworks and enforcement to protect data about children in education; (2) Introduce a 10-point certification scheme for EdTech used in school settings; and (3) Create a trusted data-sharing infrastructure to serve children's best interests and the public interest.

The United Nations Roadmap for Digital Cooperation (United Nations, 2020) invites countries to undertake a concerted global effort to encourage and invest in the creation of **digital public goods**: open-source software, open data, open AI models, open standards and open content. These digital public goods should adhere to privacy and other applicable laws and best practices, do no harm and help the attainment of the SDGs. The UNESCO International Commission on the Futures of Education (2021) considers that the best strategy for directing digital transformation towards supporting education as a common good is to ensure its democratization within a robust public sphere, which can be complemented by new EdTech entrants that aim to support various aspects of digital education. The Commission considers that the continued development of digital technologies in education in directions guided by sustainability, justice and inclusion will require action from governments, support from civil society and a broad public commitment to treating education as a space for public investment in a sustainable, just and peaceful future (Ibid., p. 112).

Cross-border data flows and national sovereignty

With specific reference to the education sector, as yet there is no comprehensive taxonomy classifying different types of cross-border education data flow. Organizations such as the UIS and the OECD, as well as ministries of education and national statistics authorities develop their own classifications systems or frameworks to analyse cross-border flows for specific objectives, stakeholders and requirements – for example, to support SDG progress monitoring through regional and global education indicators, participate in international learning assessments or analyse international student mobility and educational exchanges.

Box 11. Education Data Digital Safeguards for the public good

There is no comprehensive assessment and monitoring of the EdTech market at national or international levels needed to build trust and give peace of mind to education systems globally (Hillman, 2022). Education Data Digital Safeguards (EDDS) has been at the forefront with research and stakeholder engagement to advocate for the regulation, supervision and certification of the EdTech sector in an independent, systematic and robust manner as schools' dependence on EdTech and advancing algorithmic systems continues to grow. The dependence on proprietary systems cannot be ignored. To protect children's rights to education, these systems should be subject to systematic independent audits and licensed to operate. To mitigate these growing problems, EDDS developed the first comprehensive Quality Evaluation Framework for EdTech, together with Edtech Impact, Education Alliance Finland, WiKIT Norway and WhatWorked Education, which launched globally to serve all schools and teachers by auditing and certifying EdTech. A unifying assessment that certifies EdTech to build a transparent ecosystem of trust. Over 2,000 EdTech products are reviewed on Edtech Impact by teachers, with 400,000 unique visitors. The Framework evaluates EdTech on four pillars: lawful, human rights-respecting, ethical and safe; pedagogic value; user experience; and researched impact. The Framework ascertains that EdTech meet minimum appropriate standards and requirements; brings all stakeholders – policy-makers, teachers, students, EdTech providers – on one platform, where all requirements are transparent, verifiable and measurable.

Even without an agreed taxonomy within the education sector, it is still possible to identify various categories of cross-border education data flows generating significant educational, social, economic and cultural value. These include data shared through:

→ digital communication between multiple national ministries of education and development agencies within multilateral frameworks (United Nations, G7, G20, ASEAN, Economic Community of West African States, etc.)

→ country exchanges of information as a result of bilateral cooperation

→ participation in international and regional education conferences

→ participation in regional and international education assessments (including access to cloud-based databases shared by different countries)

→ use of online learning platforms and apps leading to the transmission of information on learners and learning contents across frontiers

→ student mobility and exchange programmes

→ agreements arrived at between countries on regional qualifications frameworks

→ Internet-based and digitized skills credentials

→ digitized cross-border sharing of information and research findings between dedicated sectoral research and development institutes, higher education and governments

→ any other education instruments which generate the exchange of education data across frontiers

Data flows across national borders when countries agree to participate in standardized international assessments and surveys, such as the OECD Programme for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMSS) or the Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (Latin American Laboratory for the Assessment of the Quality of Education) in South America. This data usually includes information on student enrolment, performance, retention and dropout rates, school characteristics, teacher recruitment and broader educational policies. The general purpose is to facilitate international benchmarking against global education standards, and to assess progress towards globally agreed goals in relation to access, equity and gender equality, in line with SDG4 of the SDGs.

LMS, Massive Open Online Courses (MOOCs), educational apps and open educational resources subsequently generate huge cross-border flows of information on learners, including student enrolment data, course progress, assessments and the sharing of academic records. These platforms have also created portals for sharing a rich array of lesson plans, digital teaching materials and multimedia resources, pedagogical approaches and best practices across countries. Online teaching communities, meanwhile, facilitate information exchange, peer support and collaboration between educators and teachers in different countries, all of which contribute to cross-border data flows.

Box 12. Aprende.org: A global digital public good at scale

Aprende.org from the Carlos Slim Foundation provides a valuable example of a scalable global digital public good, used by 25 million learners and teachers from 194 countries. The platform provides free access to a wide spectrum of content for education and lifelong learning, including materials produced specifically for the platform as well as courses from Khan Academy and leading universities such as MIT. It includes 6,900 online video courses (on mathematics, biology, the humanities, computer science and other subject areas) for primary, secondary and tertiary education. In "Training for Jobs", the platform offers 270 courses in many trades, including information technologies, construction, services, mechanics, carpentry, accounting and more. This part of the platform is directed to people needing training to open a business or retraining for those employed wishing to increase their capacities. Moreover, Aprende.org provides certification upon completion of all courses, which increases the platform's ability to support reskilling and upskilling for the world of work during a lifelong learning journey.

Under the right conditions, the benefits of cross-border data flows within a lifelong learning perspective are experienced on a shorter-term horizon. Not only does frontierless online learning expand the scope of educational opportunities, but it also fosters a more culturally diverse and inclusive education environment. This enables students and educators from different cultural backgrounds to connect and access a broad range of educational resources from various countries, including online courses, digital libraries and learning materials.

As young people and adults access online learning in greater numbers and pursue learning throughout their careers and lifetimes in formal, non-formal and informal settings, education authorities and technology companies have recognized the importance to learners being able to showcase their skills and re/connect with different levels of education, specialized training and labour market insertion programming. Digitized credentials are also gaining popularity within regional qualifications and equivalency frameworks (see Box 11).

Box 13. The European Qualifications Passports for Refugees and the UNESCO Qualifications Passports for Refugees and Vulnerable Migrants

The European Qualifications Passport for Refugees is a standardized framework and document that explains the qualifications a refugee is likely to have, based on available evidence (Council of Europe, 2023). Although the EQPR does not constitute a formal recognition act, it summarizes and presents available information on the applicant's educational level, work experience and language proficiency. The EQPR is a Council of Europe initiative. UNESCO has expanded this initiative internationally through its UNESCO Qualifications Passport for Refugees and Vulnerable Migrants (UQP) (UNESCO, 2020).

For example, academic banks of credits have been established as part of the **European Union's Erasmus+ programme** to facilitate joint degrees and student mobility, allowing the accumulation and transfer of credits between countries and programmes.

Given that data cross borders and feed into countries' education systems, the economy and society can be leveraged to generate significant private, economic and societal value. At the same time, the general and specific benefits of cross-border data flows must be balanced with consideration of legal and ethical concerns connected to the safeguarding of individuals' data privacy and the security of sensitive data, as well as other as other human rights and cybersecurity threats. Moreover, respecting cultural sensitivities is essential to maintaining trust in cross-border data flows.

In this regard, and across all sectors, emerging legal and ethical issues of concern to all countries include:

→ **Data sovereignty and control**
Cross-border data flows can result in countries' loss of control over sensitive education data, as the data may be stored and processed in foreign jurisdictions. This raises questions about who has access to the data, how the data are used and whether the data are subject to foreign regulations or legal processes.

→ **Data protection, security and surveillance risks**
Data protection laws, privacy regulations and IPR can vary between countries, which may impact the transfer of personal data, storage and processing of education data. Concerns about data privacy and security are especially raised where robust systems have not been enforced (or created) to limit who has the right to access, use and store sensitive education data obtained through EdTech and apps, LMS, cloud-based data management and transnational learning assessments. There can be risks of unauthorized access, undue surveillance, misuse of personal information and data breaches.

→ **Inequality and data bias**
Data-*driven* educational systems heavily rely on the availability and quality of data, which can be limited in certain regions or marginalized communities. Where the systems are shared across borders, biases and inequalities can be perpetuated in educational opportunities and outcomes.

→ **Cybersecurity risks**
Digital platforms and tools for facilitating cross-border data flows expose educational institutions and systems to cybersecurity threats, including hacking, data breaches and ransomware attacks. Robust cybersecurity measures and protocols are necessary to protect sensitive education data.

> Different understandings of key data terms and approaches can hold back multilateralism and country agreements on what types of data (including education data) might be tracked across borders for different purposes supporting the common good.

While there is a sense of urgency in arriving at solutions for all these issues, as highlighted in the *Digital Economy Report 2021* (UNCTAD, 2021), the global landscape for the governance of data is still fragmented, representing an asymmetry of sovereignty, with countries and regions adopting different policy approaches to regulate and safeguard data flows across their borders.

Box 14. Unlocking the potential of cross-border education data flows: Building trust and expanding access

Cross-border flows of education data empower global collaboration and knowledge sharing. The African Virtual University (AVU) serves as a compelling case study, highlighting the importance of clear frameworks for governing these data flows. AVU is a pan-African institution that has established the largest distance and eLearning network across 27 countries in Sub-Saharan Africa. It offers 219 open educational modules in multiple languages and reaches 142 countries, with significant user engagement from Brazil and the United States.

Seamless exchange of education data across borders enables AVU's collaboration with multiple universities, enhancing educational access and quality in underserved regions. For example, AVU's partnership with the University of South Africa (UNISA) allows students to access UNISA's extensive course offerings through AVU's online platform. AVU faces the complexity of adhering to varying regulatory frameworks and legal requirements across partner institutions. For instance, the African Union could lead initiatives to harmonize data protection laws and facilitate information sharing among member states. Cross-border education data flows hold immense potential for expanding access to quality education, but complying with data protection laws is crucial.

Transparency, explicability and trustworthiness

The fair use of personal data is one of the most critical issues for shaping a sustainable and prosperous digital society. Personal data have significant social, economic and practical value. Such data hold the key to improving a range of services and products provided by governments, companies and organizations. Today, storing personal data can be seen as a liability, while having permission to use a piece of datum is an asset. EdTech companies that grasp this early enough and build ethics and transparency into the very design of their products and services, are in a better position in a landscape with increasing data regulation. In principle, personal data-based services should be built on mutual trust, but examples of this trust and transparency in EdTech services are not the norm at present.

Building trust in data-*driven* decision-making is a complex process, requiring transparency, explicability and quality in datasets and algorithmic applications. **Transparency** pertains to data knowledge, including ownership and consent, encompassing how, when, why, where and by whom data was collected, processed and used. Transparency around data production and use can build trust in the applications of data-reliant tools. As such, trustworthy data – and trustworthy collectors of data – are crucial for data-*informed* decision-making.

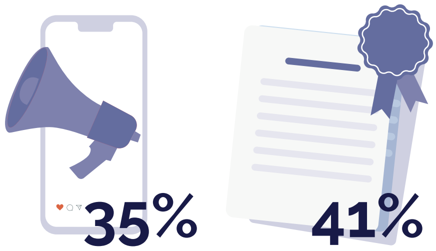
> **Educational institutions should work to assure individuals own and control their personal data, and, in the case of children, families should be actively involved in decision-making. When possible, learners should be able to "opt-out" of data capture and still retain full access to educational opportunities.**
> (UNESCO and Dubai Cares, 2021, p. 6)

On the African continent, the expansion of the EdTech industry has been significant, with an estimated US$57 billion estimated on EdTech expenditure by 2030. In the face of this rapid growth, many African countries are developing online protection frameworks specifically directed at the protection of children's data, such as the Nigerian Data Protection Regulation. This legal instrument contains a specific provision on privacy policy to build trust between end-users and EdTech providers, for it requires platforms to display their data policies prominently on their websites in simple language (Odhiambo *et al.*, 2021).

However, transparency and trust are not enough. Especially in increasingly complex AI-driven technologies, **explicability** is a key component to build trust in data practices. Explicability concerns the understanding of data-driven systems, their inputs, functioning and contributions to outcomes, which necessitates traceable and comprehensible subprocesses. Trustworthiness also deeply entails the concept of data quality. Data could be considered high-quality "if they are fit for their intended uses in operations, decision making and planning" (Kerr *et al.,* 2007). Alternatively, the data are deemed of high-quality if they correctly represent the real-world construct they refer to, which includes the use of proxies.

There is a clear need to improve explicability, transparency and quality in data-reliant tools to build trustworthiness in data practices in education and beyond. In the absence of a reliable source to turn to, individuals find themselves adrift, uncertain about where to find credible information and whom to rely upon. This pervasive global infodemic has eroded trust in the legacy information sources to unprecedented levels, with social media (35 per cent) and owned media (41 per cent) ranking as the least dependable sources. Traditional media (53 per cent), once regarded with higher confidence, has witnessed an alarming 8 percentage point decline in trust on a global scale (Edelman, n.d.).

This pervasive global infodemic has eroded trust in the legacy information sources to unprecedented levels, with...



While addressing this issue, it becomes evident that enhancing the elucidation, openness and excellence of data-*fuelled tools* is imperative to establish credibility in data practices, both within the realm of education and beyond. Given the pervasive deficiency in trustworthiness across diverse regional contexts, it becomes paramount to furnish data consumers with a universal annotation system that signifies the reliability quotient of input data whenever it undergoes transmission. This task poses significant challenges, particularly in scenarios where substantial data volumes are incessantly produced and disseminated throughout the system.

Being able to do so is particularly challenging, especially when around 2.5 exabytes of data are generated each day and continuously transmitted across the system (Domo, 2017). Further, solutions for increasing trustworthiness of the data, like those to specifically target the data quality, may be expensive and may require access to data sources that have access restrictions due to data sensitivity (Bertino and Lim, 2010).

From an inverse perspective, effective use of data can increase trust and accountability in the education sector as a whole. For example, **Pakistan** and **Nepal** have initiated targeted efforts to enhance transparency and accountability in the education sector by using data. In Pakistan, the school report card initiative, a part of the Punjab Education Reforms Roadmap, was launched to collect and share school performance data, including student attendance, teacher presence and infrastructure conditions, through a public web portal. Parents, communities and policy-makers gained the ability to hold schools accountable and make informed decisions. Based on a study from Harvard University, implementation of the school report card initiative yielded remarkable results through improved educational outcomes (Chaudry and Tajwar, 2021). These schools witnessed higher student test scores, increased enrolment rates and reduced teacher absenteeism. Parents and communities were empowered to monitor and advocate for necessary improvements, while policy-makers effectively allocated resources to address gaps.

Box 15. Connecticut General Assembly enforces transparency and explicability in student data use

In 2016, Connecticut's General Assembly passed Public Act No. 16-189, the US state's first data privacy law for students. The statute's requirements are commonsensical – protections against targeted advertising and unauthorized sharing, destruction of data upon the termination of contracts, etc. – brought forth by parents asking for transparency into and control over the data concerning their children. Lawmakers likely intended the burden of compliance to fall on the EdTech providers with which schools shared student records and information, but that is not what happened.

Instead, school leaders in 200 local education agencies had a short window during which to review the terms of each EdTech app that they used, with an average of 1,400 unique titles per agency. This process imposed a heavy administrative burden on school leaders and often teachers, where governance decisions fall as to which products to use based on local instructional needs. To introduce economies of scale, the Connecticut Commission for Educational Technology created a statewide library of educational software. Listings in the library appear only after EdTech companies have digitally signed assurances that their terms of use and data privacy agreements fully comply with Connecticut's law. Introduction of the library shifted the onus of review and compliance back onto providers, and clarified for EdTech companies the often confusing interpretations of the law they received from different education leaders – individuals skilled at running schools rather than in contracting and privacy. To date, this approach has saved approximately US$10 million and 40,000 staff hours that otherwise would have been spent on compliance activities.

Inclusion, diversity and fairness

Data are generated, processed, cleaned, analysed, preserved or destroyed by humans. Algorithms that are trained on these datasets are likewise driven by human decision-making. In short, both the datasets and the systems built atop these datasets are subject to human bias. The international community is becoming aware of this reality, with scholars and experts warning against blindly trusting the supposed neutrality of data. Similar multilateral understandings need to be built and developed regarding data for learning ecosystems, too.

All algorithms are prone to some degree of error. In large EdTech companies, even a tiny error can have a large impact. The impact could be even more detrimental if algorithms are applied within education in an opaque, ill-designed data architecture. Every effort should be made to audit such systems for **fairness**, and ensure the trade-offs between flexibility and efficiency are transparent and treat individuals with compassion and respect.

> **Be cognizant that AI applications can impose different kinds of bias that are inherent in the data on which the technology is trained and which it uses as input, as well as in the way that the processes and algorithms are constructed and used.**
> (UNESCO, 2019)

At the institutional level, practices that automate grading systems or employ predictive analytics without disclosure have been met with substantial criticism. One of the most publicized instances was in the **United Kingdom**, where an A-level grading algorithm was used to assign grades to A-level students during pandemic school closures (Berridge, 2020). The results of A-level exams for students aged 16–18 directly affect the higher education opportunities available to them. When the Government asked teachers to assign A-level grades in place of exam results, there were more higher grades than usual. Therefore, in an attempt to objectively standardize the grades of all students, the United Kingdom Government turned to the Office of Qualifications and Examinations Regulation (Ofqual), which used an A-level grading algorithm to assign grades to A-level students (Ibid.).

The result saw 40 per cent of students receiving lower grades from the Ofqual grading algorithm than teacher assessments had indicated they would receive. Furthermore, there were striking trends between students who received lower grades compared to students whose grades remained on par with teacher assessments or even improved upon them. Many more students from state schools saw their grades cut, therefore limiting their prospective higher education opportunities, while students from independent and private schools saw their grades improving (Allegretti, 2020). The results caused an outcry in the UK and sparked protests up and down the country (Castle, 2020). In response, the Government reversed their decision to put A-level assessments in the hands of the Ofqual grading algorithm and fell back on teacher assessments whenever they were higher than those awarded by the algorithm (Ofqual, 2020).To prevent undue trust in algorithmic decision-making over that of human teachers, the education sector can learn from common misuse in other sectors. Organizations increasingly rely on algorithms to help make decisions that impact people's lives, including who gets a bank loan (Unitas Financial Services, n.d.), a job (Metz, 2020) or jail time (Zhu, 2020). Public backlash has led to proposals like the Algorithmic Accountability Act (Jones Day, 2019), which would require the United States Government to develop rules that mitigate algorithmic bias and provide ways for citizens to appeal automated decisions.

Regarding **inclusion** and **diversity**, at the local level, Member States should promote equitable data representation between rural and urban areas, and among all people regardless of ethnicity, gender, age, language, religion, political opinion, national origin, social origin, economic or social condition of birth, or disability and any other grounds, in terms of access to and participation in the data-*driven* system lifecycle (UNESCO, 2023c).

Box 16. ProUni: Empowering inclusion and diversity in Brazil's higher education

Brazil's Programa Universidade para Todos (ProUni) is a remarkable initiative promoting inclusion by providing scholarships to economically disadvantaged students. Since its establishment in 2004, ProUni has granted over 2.5 million scholarships for undergraduate courses at private higher education institutions in Brazil (Schneider *et al*., 2019). Education data governance in Brazil plays a vital role in ensuring fairness and addressing disparities. By evaluating programme outcomes and analysing demographic data, policy-makers can identify areas for improvement. This data-driven approach enables targeted interventions and adjustments to enhance inclusion and fairness within ProUni. Studies show that ProUni has significantly increased enrolment of students from lower socio-economic backgrounds in private higher education institutions (Bruns *et al*., 2011; Garcia, 2012; Lépine, 2016). By examining the distribution of scholarships across different regions and demographic groups, education policy-makers strive to reduce educational inequalities and provide opportunities to a diverse range of students.

Managing personal data and safeguarding student privacy are paramount. Brazil's General Data Protection Law (LGPD), enacted in 2018, is a leading example. It ensures transparency, consent, purpose limitation, data minimization and accountability in educational institutions (Brazil, General Personal Data Protection Law).

- **Transparency:** The LGPD mandates clear information on data collection and processing, thus fostering trust.
- **Consent:** Institutions must obtain explicit and informed consent from students or guardians.• **Purpose limitation:** Personal data should be used only for authorized purposes.
  - **Data minimization:** Data minimization emphasizes collecting only necessary information.
  - **Accountability:** Robust security measures and accountability are required.

The LGPD enables Brazilian educational institutions to comply with data protection laws and protect student privacy. It sets international best practices for responsible data management and upholding students' rights.

At the international level, the most technologically advanced countries have a responsibility to show solidarity with the least advanced to ensure that the benefits of data-*driven* technologies are shared such that access to and participation in the data-*driven* system lifecycle for the latter contributes to a fairer world order regarding information, communication, culture, education, research, and socio-economic and political stability.

It is well known that digital proctoring methods, such as tracking eye movements, are commonly operationalized by for-profit technology companies to measure engagement with products and advertising. What happens if we define engagement in terms of eye movement in learning spaces? Some EdTech companies have already attempted to quantify traditionally qualitative data on social and emotional factors by deploying AI facial coding algorithms that seem to identify facial expressions of emotion. When evaluating such data for a learning model according to the United Nations Convention on the Rights of the Child (United Nations, 1989), significant ethical, legal and socioemotional risks were revealed (McStay, 2020).

Reliability, accuracy and completeness

The heavy cost and capacity requirements of data collection, cleaning and analysis – combined with the limited ability of education data to capture the full extent of learning outcomes – prevents some contexts from developing a data culture that would improve general trust in the reliability and usefulness of data due to higher data literacy and exposure in education settings. The complexity of assessing and **auditing the reliability and accuracy of data analysis and modelling** for insights is therefore heavily dependent on context. In systems where learning data are readily available, these data represent only limited aspects of the past, yet are often used to predict the future. This can limit outcomes and individual development if not done properly and well. Moreover, the assessment of degree of impact on a learner's trajectory is also contextually dependent, such as the difference between recommending one course as opposed to another through predictive analytics versus inaccurately grading a child in a consequential exam. Auditing tools would need to be responsive to these context differences when assessing potential data misuse.In this context, some of the main barriers to unlocking the beneficial potential of data for learning relates to the availability and quality of the data, which can undermine accuracy, completeness and reliability. According to the UNESCO *Global Education Monitoring Report 2017/8* (UNESCO, 2017), around 75 countries have incomplete or unreliable data on learning outcomes, hindering their ability to monitor education systems effectively. It is important, therefore, to consider how to enhance the analysis of data that may typically be messy, restricted or incomplete. On the other hand, the pressure to collect as much data as possible to present a more holistic picture of student progress is not without its risks. As evidenced by the study in Box 14, pressures to produce complete and reliable data can create a culture of mistrust in teacher judgment, pressuring pedagogical proof in all aspects of student learning.

Box 17. Case study: Pressure for "infinite data collection" seen as detrimental to student learning and teacher trust in Queensland, Australia

A 2021 study of teachers in Queensland, Australia (Daliri-Ngametua *et al.,* 2021) revealed a perception that teachers were constantly engaged in collecting student data to justify their teaching practices with "hard evidence" of student growth (Ibid., p. 7). This created a culture of mistrust in teachers' professional judgment from both school leaders and parents. Schools increased auditing and surveillance, and parents pressured teachers to provide them with material and measurable data to prove claims about learner performance, classroom behaviour and teaching practices.

The volume of data that Queensland teachers were expected to collect was perceived as "detrimental to student learning because it took away the time and energy needed to focus on providing substantive learning opportunities to students" (Ibid., p. 9). As a consequence, teaching became less about "doing" and more about performing progress, which, in turn, diminished the power of teachers' professional judgment and strengthened the sanctity of explicit and quantifiable student learning metrics, which reflect only a partial view of learning experiences and outcomes (Hardy, 2021).

Privacy, consent, safety and security

Privacy

Privacy has become possibly the most obvious concern relating to the collection of all types of personal data. Data can be compromised, exposing them to a plethora of risks and threats, such as identity theft and blackmail. Within education, however, there are further issues to consider. For example, sensitive dimensions of student identity, such as family income, special needs, counselling files, grades, addresses and contact information, can be uncovered if a school data management system is compromised. Even detailed aspects of students' socio-economic backgrounds can become visible if students participate in video classes that can put the interiors of their homes on display, and from this, companies could accrue potentially marketable information.

Anonymizing and pseudonymizing student data at the individual level does not go far enough to protect learner privacy. If the security of the system is weak, then individual data can easily be triangulated by a hacker who combines multiple databases to enable the reidentification of individuals (Quinton and Reynolds, 2018). There are three main techniques for de-pseudonymizing data, which can all be used in tandem in attempts to de-anonymize data: combing datasets, pseudonym reversal and inadequate de-identification in the first place. Attempts to de-anonymize data can identify the real identities of individuals using direct identifiers.

Moreover, the integrity of data used for monitoring and credentialing processes can itself be called into question, as learner identity needs to be secure enough for digital credentials to be trustworthy. Important questions remain, then, not just concerning the privacy and anonymity of learners, but also relating to the legitimacy of data for learning as a tool. The organization EdSAFE AI Alliance is attempting to address the application of data rights in education spaces, and is leading an international effort to develop benchmarks, standards and certifications to establish trust for the use of AI tools in education (EdSAFE AI Alliance, n.d.).

Consent

Another tension that stems from the protection of learners' data is the issue of meaningful consent. Although providing consent is currently the main feature of most personal data protection efforts, enabling meaningful consent remains one of the most difficult challenges in the digital context. It is possible to request consent using both opt-in and opt-out techniques. However, data use policies are often very long, hard to find and difficult to understand. In the face of opaque techniques, consent may reflect only the need to access educational services, rather than a true acceptance of the terms presented in the privacy policy. Meaningful consent, on the other hand, is a policy that is understandable for all learner types and is age-appropriate.

How do we ensure that learners are providing genuine consent that does not require them to read pages of dense text, and that is easily understood? How do we understand the temporality of consent? Who expresses the minor's consent – the minor or their parents? If a learner consents to their data being used at one specific moment, does that also mean they have consented to its reuse in the future or to its inclusion in a larger dataset, the use for which was not disclosed at the moment the learner consented? Such questions are challenging to answer with a strong legal basis, which adds to the complexity of regulating education data use in practice.

Safety and security

When learner data are stored online, it becomes much easier for malicious actors to access and exploit, and, increasingly, breaches of this nature can expose the data of huge numbers of learners all at once. The question is, then, how do we prevent and mitigate data breaches? There are efforts to safeguard learners at both the local and international levels. At the local or school level, learner data should be protected from exploitation, but also from the threat of security breaches and school cyberattacks.

At the system level, there is some debate as to whether these protections are ensured by existing international standards that apply to all individuals, such as the GDPR, or whether a specific international normative framework for the learning context is needed. If such a normative framework were developed to safeguard learner data, then it should include the voices of teachers, schools and local agencies, whose efforts should not be sidetracked by the introduction of new regulatory instruments.

Two such frameworks relevant to education uses cases are the Global Education Privacy Standard (GEPS) and the Global Education Security Standard (GESS).

GEPS and GESS provide essential guidelines for data retention, deletion and disposal, ensuring proper data management throughout its lifecycle. They emphasize the importance of detailed recipient information for issue resolution, including data handlers and affected countries, during data transfers. As global standards for education privacy and security, GEPS and GESS ensure student data are handled with the utmost care, maintaining privacy and security amidst the evolving digital education landscape.

See Table 6 for more information on the two frameworks.

Box 18. Finland's MyData Global initiative

MyData Global, funded by Finland's Ministry of Transport and Communications is an umbrella term for a human-centric approach to personal data. The core idea is that individuals should be in control of data about them. The MyData approach aims to strengthen digital human rights, while also opening up new opportunities for individuals to access the practical tools needed to exercise them. MyData is an alternative vision that offers guiding technical principles for how we as individuals can have more control over the data trails we leave behind in our everyday actions. The main idea is that we should have an easier way to see where our personal data go, specify who can use the data and alter these decisions over time. Legislation, regulation and technological changes can all contribute to the realization of MyData. Education systems should pay close attention to the evolution of this initiative and how they, too, can contribute to a realization of this vision.

Table 6. Global Education Privacy Standard (GEPS) and Global Education Security Standard (GESS)

GEPS	GESS
The Global Education Privacy Standard (GEPS)[13] is a framework dedicated to safeguarding student privacy and ensuring data security in the education sector. With data breaches becoming increasingly prevalent, GEPS goes beyond initial data privacy agreements by establishing automatic communication between educational institutions and technology providers. GEPS offers clear guidelines and restrictions on data usage, defining permissible and prohibited actions to create boundaries for handling student data. It prioritizes transparency and accountability by stipulating the purpose for which data are provided. Furthermore, GEPS sets legal obligations and technical benchmarks, ensuring compliance with relevant laws, regulations and industry best practices. This comprehensive framework introduces a "green list" that authorizes access to specific data elements for educational purposes while maintaining privacy. It also enables the segmentation of data into subsets based on factors such as senior years of education, minimizing unauthorized or unnecessary data usage.	To address student data privacy obligations, the Student Data Privacy Consortium[14] has successfully fostered collaboration between EdTech providers and school districts. However, the lack of mandated security requirements for providers poses challenges in certifying privacy and security compliance. To bridge this gap, the Global Education Security Standard (GESS) Project Team has developed a matrix incorporating existing security frameworks. The aim is to identify a core set of controls applicable to PK-20 data (Murray, 2023), covering the entire education spectrum. Seeking feedback from industry experts, the project team intends to refine and guide their work further. Version 1.0 of GESS integrates controls from renowned security frameworks such the United States National Institute of Standards and Technology's NIST SP 800-171 and NIST SP 800-53,[15] the US Center for Internet Security's CSI Critical Security Controls version 8,[16] the United Kingdom National Cyber Security Centre's Cyber Essentials,[17] the Australian Cyber Security Centre's Information Security Manual,[18] the New Zealand Government Communications Security Bureau's New Zealand Information Security Manual[19] and the Australian Safer Technologies 4 Schools.[20] By combining these controls, GESS establishes comprehensive and unified standards to ensure education data security.

The challenge of providing safety and security for learner data is deepened by socio-economic disparities, since the maturity of protections aligns with a country's economic development. Of the 76 countries reviewed by a recent Massachusetts Institute of Technology and Infosys study on regulatory frameworks for the use of cloud models, two low-income countries ranked seventy-first (Uganda) and seventy-sixth (Ethiopia). Lower-middle-income countries did not rank higher than forty-third (MIT Technology Review, 2022). This puts students in less wealthy areas at a disadvantage.

Even though there are few case studies that show promise, security breaches are becoming increasingly common. A McKinsey Global Institute (2021) report outlines that:

13 See https://privacy.a4l.org/geps/.
14 See https://sdpc.a4l.org/gess/gess_document.pdf.
15 See https://sdpc.a4l.org/gess/gess_attributions.php.
16 See https://www.cisecurity.org/.
17 See https://www.ncsc.gov.uk/cyberessentials/overview.
18 See https://www.cyber.gov.au/.
19 See https://nzism.gcsb.govt.nz/.
20 See https://st4s.edu.au/.

Breaches can occur during transfer of data, or at any institution involved in the open data ecosystem, such as a bank or fintech. For example, when data transfer is achieved via APIs, a hacker who breaches such an API can hijack any apps that use the interface to collect data.

In the context of learning, this means that any data passing through a student information system or a digital learning programme faces the threat of data compromise. Naturally, with the use of digital platforms and programmes growing globally during the period of COVID-19 school disruptions, the number of cyberattacks and data breaches has also increased (Levin, 2021). For example, the publishing company Pearson, known for its textbooks, was fined US$1 million to settle charges that it had "misled investors about a 2018 cyber intrusion involving the theft of millions of student records, including dates of births and email addresses, and had inadequate disclosure controls and procedures" (United States Securities and Exchange Commission, 2021).

Many local, regional and national education bureaux simply do not have the human or capital resources to do this in a best practice and sustained manner. Public schools may have fewer resources than private corporations to train their students and staff, and to configure and secure school networks and devices to prevent online disruptions. As such, ensuring cyber protections is an equity issue, as learning environments with fewer resources are more vulnerable to attack.

Box 19. Illuminate Education data breach in the United States

A data breach that occurred during a cyberattack on Illuminate Education in the United States in January 2022 is so far known to have affected nearly 2 million students across the country. The breach has affected learners and educators in five different states, with over 820,000 students impacted in New York City alone (Kuykendall, 2022).

Officials from some of the affected school districts have stated that data included in the breach include learners' names, dates of birth, races or ethnicities, and test scores. Representatives from at least one of the districts indicated that the breach also included particularly sensitive information, such as behavioural records, tardiness rates, information about disabilities and migrant status (Singer, 2022). Breaches of this nature can have long-lasting consequences for the affected learners, with sensitive and confidential information becoming available for public scrutiny.

Unfortunately, the full extent of the Illuminate Education data breach is yet to be known, as the company working with the stated aim of assisting education partners and educators reach new levels of student performance through the utilization of data, claims on its website to serve over 5,000 schools nationwide, covering a total of over 17 million enrolled students in the United States.

Financial sustainability

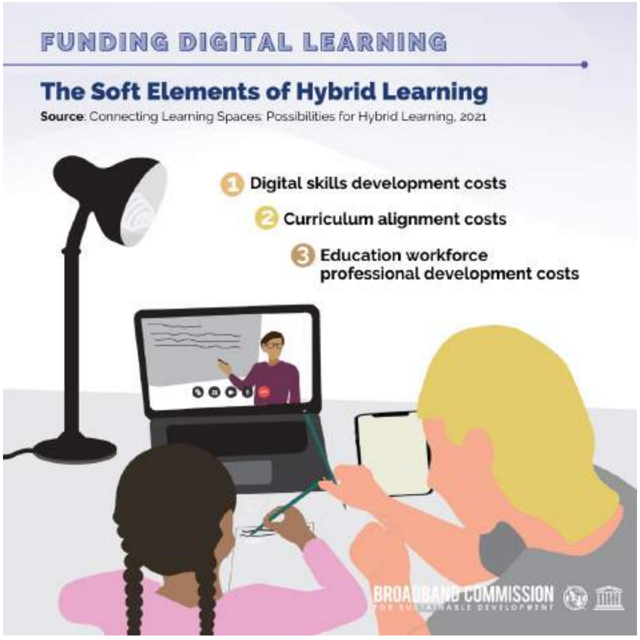
The costs and consequences of expanding learning data models are key challenges to consider when weighing the potential benefits of these systems against the potential risks to environmental well-being. In parallel with concerns about data architecture, there are great concerns about the deployment and integration of new data practices where these must be incorporated into already under-resourced education systems. The adoption of emerging data practices would require major investments in hardware, software and human resources to support the shift away from centrally managed data architecture towards more vertically and horizontally integrated management systems.

Furthermore, the adoption of certain frontier technologies raises additional issues for financing and planning departments. Even if countries could raise the funds, investments are often a complex set of moving targets that generate cumulative costs and require complicated forecasting and financing skills. A key feature of cloud to edge technology, for example, is the opaqueness of the costs involved. Cloud providers seemingly make it intentionally difficult to understand costs, meaning that a whole industry has recently emerged to help companies understand and compare cloud storage costs.

Sustainably financing data for learning ecosystems, therefore, is both vital to ensuring associated challenges and risks are mitigated efficiently, as well as difficult due to the technical solutions offered and their increasingly complex implementations. This is before taking into account the specific elements that need to be paid for to support data for learning systems, what the associated costs may be, and how to source and engineer the required financing and funding.

As described in the Broadband Commission's (2021*b*) report on digital and hybrid learning, financing digital learning comprises both "hard" and "soft" elements, many of which are similar to the costs of financing data for learning.

Figure 9. Hard and soft elements of funding digital and hybrid learning



Source: Developed for the Broadband Commission (2021*a*) report Connecting Learning Spaces: Possibilities for Hybrid Learning.

Hard elements include configuration and data storage requirements, hardware costs in the data hub, capital and annual running costs, energy consumption and cooling water consumption requirements for data centres. Soft elements include database management, cloud-based computing, and training for data literacy and advanced data analytics. Additionally, how and where the software package was developed, either in-house or purchased from a vendor, influence costing. If developed in-house, the amount of time expended on development is a factor in the cost calculation, as is the administrative structure for data systems – in other words, the maintenance and IT staff needed to keep the hardware and software operational. Thus, the human capacity necessary to maintain data-rich systems comes with a large financial cost.

The adoption of certain frontier technologies raises additional issues for financing and planning departments. Even if countries could raise the funds, investments are often a complex set of moving targets that generate cumulative costs and require complicated forecasting and financing skills. Cloud providers seemingly make it intentionally difficult to understand costs, meaning that an entire industry has emerged to help companies understand and compare cloud storage costs. For example, cloud storage costs are derived from a complicated formula based on: (1) the number of gigabytes stored; (2) how frequently the data are accessed and retrieved; (3) network bandwidth; (4) copy costs across multiple locations, such as US, Europe and Asia (particularly relevant for international enterprises); and (5) disaster recovery to move from on-premises storage to cloud and vice versa (Veritas Technologies, n.d.). All these factors accrue because there is a per gigabyte cost each time servers in different domains communicate with each other, and another per gigabyte cost to transfer data over the Internet.

The cost of data for learning

Yao *et al.* (2021) examined the costs associated with financing universal digital learning, detailing different costing elements and drivers and the estimated costs associated with them. Of all the different costing elements, the policy brief identifies the provision of affordable data as the most expensive, with a price of US$498 billion for the period 2021–2030.

- → US$410 billion for universal electricity
- → US$428 billion for universal Internet connectivity
- → US$498 billion for making data usage affordable (zero-rating)
- → US$46 billion for the recurrent delivery of digital learning

However, with many low- and middle-income countries (LMICs) in need of reliable electricity supply and Internet connectivity in many regions, it is difficult to disassociate the costs of connectivity and infrastructure enabling factors, or the recurrent delivery costs, such as the provision of devices, digital learning and capacities development, when examining the sustainable financing of data for learning. Also, the UNICEF analysis includes conservative estimates on certain pricing elements assuming, for example, a sample device cost of US$20 per learner and a sample smartphone cost of US$50 per teacher.

An analysis by the World Bank (2023, forthcoming), based on an estimate of 2.5 billion young people under the age of 25 living in LMICs, shows that according to the UNICEF costing model, LMICs will have to invest roughly US$56 per learner each year to cover all associated costs relating to digital learning. Therefore, even with conservative estimations of costs driving analyses, the projections represent massive challenges for LMICs when compared against prospective education budget estimations.

A UNESCO (2023) study built upon the UNICEF model to estimate the one-time and annual operating costs of universal digital learning across three models of digitalization: basic offline; fully connected schools; and fully connected schools and homes. The findings of the model revealed that investment required for the universalization of digital education is concentrated in South Asia and Sub-Saharan Africa. Across all income groups, approximately US$1.4 trillion would be needed for capital expenditure (one-time investments), with an additional US$558 billion annually for operational expenditures (Ibid.).

> A 2023 study by UNESCO estimates that fully connecting schools in low- and lower-middle-income countries would cost US$183 billion in one-time expenditures and US$111 billion in operational expenditures.

This level of expenditure is unlikely to be sustainable for governmental budgets alone, indicating a requirement for multilateral engagement when developing and deploying sustainable financing strategies to support digital learning and appropriate data ecosystems, particularly in LMICs.

Financial engineering for long-term self-sustainability

Assessing the high upfront costs for digital technology solutions, including connectivity and infrastructure considerations, encompassing both the hard and soft elements associated, is inadequate for accounting for the actual finance and funding requirements for digital education in the long term. The total cost of ownership is an essential tool for comprehensively assessing the true costs of implementing and maintaining modern data for learning ecosystems and their associated digital learning technologies. The model considers all associated costs beyond the initial expenditure on acquisition, also incorporating operating, sustainability and upgrade costs into financial forecasting. According to the World Bank (2023, forthcoming), budgets should be prepared to cover between 50–100 per cent of acquisition costs every five to six years to account for operating and sustainability costs.

→ **Acquisition costs** may go beyond the procurement of hardware and software solutions. They could also include taxes, transportation, installation and testing, financing costs and initial teacher training.

→ **Operating costs** include all ongoing expenses, such as maintenance and support services to ensure the reliability and smooth operation of devices and systems, connectivity costs, electricity, paying for adequate security measures, insurance, ongoing teacher training, ongoing learner training, and ongoing monitoring and evaluation of the systems.

→ **Sustainability costs** account for the expenses associated with the long-term viability and upgradability of digital education systems, including the costs associated with upgrading devices and software periodically to keep up with technological advances, replacement costs for worn out or outdated equipment, and proper disposal or recycling measures when upgrading is not possible or viable.

There are many examples of financial engineering efforts to support digital learning ecosystems taking place at different scales around the world. Building on the financing framework for hybrid learning outlined in a Broadband Commission report (Broadband Commission, 2021*b*), the *Review of Alternative Country Models and Strategies for Financing Digital Learning* (Razquin *et al.,* 2023) investigated additional types of financial models that countries have used to fund digital transformation in education: universal service and access funds (USAF), zero-rating practices, national capital investments earmarked for digital learning, federal- or state-level competitive government grants, commercial models and PPPs with EdTech companies. These models offer means to implement programmes through different terms that manage the relationship between the funder and the funded.

As the UNESCO paper showed, however, the scale and complexity of connectivity, infrastructure and digital learning initiatives often demand the innovative implementation of funding models and broader combinations of multiple models to support the roll-out of larger projects.

Zero-rating access educational content was a measure adopted by many countries during the COVID-19 pandemic as part of an effort on the part of governments and telecommunication providers to ensure continuity of learning. Zero-rating refers to an arrangement where mobile network operators exempt certain data usage, typically for educational purposes, from being charged to customers. This means that users can download and upload online content without it counting towards their data allowances or incurring additional costs. Therefore, zero-rating provides free Internet access for specific content, treating users and content providers different in terms of data charges.

> **Zero-rated access to education content can allow teachers and learners to upload online content without it counting towards their mobile network operator data allowances or incurring additional costs.**

> **If governed by fair use principles and subsidized by governments, universal access funds or corporate social responsibility efforts, the model could be a step towards digital learning platforms that are free, public and open.**

In some cases, governments are considering providing incentives of various sorts for Internet providers to maintain zero-rated access to education platforms, with USAF sometimes used to subsidize such practices. In other cases, telecommunication companies are committing to continuing these practices, whether as a way to differentiate themselves in competitive local markets or as part of their corporate social responsibility efforts (or some combination of the two).

However, with zero-rating practices applying to specific activities and applications, it is worth noting that the success and impact of such an initiative rely on the continuous feedback loop between stakeholders, including students, educational institutions and the telecommunication provider. This feedback loop ensures that the initiative evolves to meet the changing needs and aspirations of the educational community it serves.

Table 7. Additional financial engineering models to support data for learning

Model	Description
USAF	USAF are public funds created by governments to improve access to telecommunication in underserved areas where it may be more difficult to establish a business case. Typically, they have been used to target connectivity, the expansion of national research and education networks, funding of ICT labs, initiating digital skills and capacities development projects, and have even been used to subsidize household purchases of devices. They are not without challenges, however, as they may come with legacy regulations outlining what they can be utilized on, and the value of the USAF pot will be different from country to county.

National capital investments	Governments in high-income countries may have developed and long-sustained dedicated budgets for digital learning solutions and initiatives. There are LMICs that have also established long-term funding support in these areas, but they often come in the form of emergency funding or manifesto promises and programmes, which can be subject to changing political winds. However, there are many examples of large-scale cases of countries using innovative methodologies for raising and deploying capital to support digital transformation in education. The use of radio spectrum auctions has long raised capital for governments, and also provides clear examples of how national resources, such as the finite radio frequency spectrum, can be leveraged to raise capital for public expenditure or even pass responsibility for enacting digital transformation onto the private sector.
Competitive government grants	Government agencies at the local, state and federal levels may offer grants for digital education initiatives. These grants may be competitive, but they can provide substantial funding. Competitive grants often come backed by large legislative bodies, which can increase the resources available to the project receiving the grant. The funding of public grants can be subject to changing political circumstances, meaning they are particularly suitable for one-time purchases or short-term funding needs.
Commercial models and PPPs with EdTech companies	Long-term contracts between a private party and a government entity for providing a public asset or service, in which the private party bears significant risk and management responsibility, and remuneration is linked to performance. These partnerships can provide funding, expertise and other resources to support the initiative. There is often significant risk attached for private sector partners regarding the long-term sustainability of the initiatives, particularly when it comes to pre-financing and governments that may lose control over initiatives.
Social impact bonds (SIBs)	SIBs are innovative financial instruments that incentivize and finance initiatives based on the results and impacts achieved (OECD, 2016). The impacts targeted by SIBs can range from improving computer skills, reaching marginalized communities and disadvantaged children, providing teacher training, enhancing literacy and numeracy skills among children, and even reducing the number of children in juvenile justice systems. SIBs aim to transfer the risk from the government and service providers to investors, offering them an opportunity to contribute to social outcomes while potentially earning financial returns.
Seed and microfinancing	Loans and small loans are provided to assist with the purchase of critical goods that meet the criteria laid out by the initiative or to cover start-up and development costs. Seed funding and microfinancing are becoming increasingly popular as ways to fund technology programmes in areas where affordability and access to funding and financing are limited.
Consortium purchasing/ Bundled service agreements	A collaborative approach that sees schools or educational institutions collaborating to purchase technology equipment and services in bulk from vendors or service providers. This approach can allow institutions and households to obtain more favourable pricing and contract terms than they could achieve individually.

Environmental sustainability

In addition to financing, the environmental costs of weaving data-*informed* learning into education spaces are often overshadowed by a conflation of digital transformation with the green transition. There are many environmental factors to consider, however, with data ecosystems and the digital and ICTs that drive them representing contributions to one of the largest greenhouse gas-emitting industries (Belkhir and Elmeligi, 2018). The vast energy expenditure of the ICT solutions and the digital technology needed to sustain data-rich educational models comes with a significant carbon footprint and numerous environmental impacts that must be considered when weighing the benefits and risks of a data for learning policy approach. For example, ICT-related operations are expected to represent up to 20 per cent of global electricity demand, with one-third stemming from data centres alone (Jones, 2018). Furthermore, more devices and increasing amounts of electronic waste (e-waste), and subsequent energy use also represent a significant challenge when considering the environmental footprint and sustainability of expanding data for learning ecosystems.

Environmental impacts to consider

Estimates from a broad range of experts and researchers claim that the ICT sector consumes between 5 and 9 per cent of the world's total electricity supply (Belkhir and Elmeligi, 2018), which comes attached to its subsequent carbon footprint. Much of this energy usage comes from data centres, cloud services and connectivity. However, going beyond carbon footprint alone, ICT solutions and digital technologies can, throughout the different stages of their lifespans, cause up to 13 different types of environmental impacts, including air pollution, greenhouse gas emissions, mineral resource depletion, forest degradation and deforestation, terrestrial pollution, freshwater pollution, marine pollution, human toxicity and waste generation (Arushayan, 2016). These stages include product design, materials extraction and processing, product manufacture, packaging and distribution, use and maintenance, and end of life and disposal or beginning of new use cycle.

Data centre energy use

Discussing the energy use and demands of data centres marks a direct correlation between data-intensive practices and their subsequent carbon footprint and environmental impacts. This means drawing a line from everyday digital activities in new and expanding data for learning ecosystems to the previously mentioned direct and representable impacts. There should be an understanding of this built into all policy decisions made in this regard, particularly in the context of the growing global demand for data-*driven* practices and the subsequent need for data centres (McKinsey and Company, 2023).

Recent estimates, however, show that efficiency improvements have helped to limit the growth in energy demand from the data centre sector (International Energy Agency, 2023). This limited growth in energy usage in the face of increased demand for data is due in part to efficiency gains found in modern ICT hardware and the trend of moving away from smaller decentralized networks of data centres to larger and more efficient hyperscale data centre facilities (Ibid.). This shift, however, only further highlights the importance of developing suitable data governance frameworks incorporating cross-border data flows. At the least, best practices from these more efficient facilities should be prioritized if data centre infrastructure is to be developed and enlarged to support expanding data for learning ecosystems.

Devices and e-waste: Another double-edged sword

A shift towards smaller devices, such as smartphones and flat panel screens, and away from larger personal computing devices and CRT screens, could significantly impact the associated carbon footprint of digital technologies and the data-based ecosystems they facilitate (Forti *et al.,* 2021). However, although devices may become smaller, increasing numbers of them will result in an increase in associated environmental issues if more sustainable production and redistribution practices are not seen as a priority.

As mentioned earlier, ICT devices come with a host of associated environmental impacts across their lifecycle, as well as poor human rights records in many locations where the required natural resources are being extracted (Merk *et al.,* 2021). Accordingly, higher levels of e-waste are another challenge that comes with increased proliferation of the devices required to support the expansion of data-based ecosystems. Beyond the considerable amount of e-waste associated with the development of new devices (Vereecken *et al.,* 2010), old devices that have reached the end of their lifecycle also represent a growing challenge globally. While advances in efficiency and bolder commitments to environmental sustainability may affect current predictions, the numbers reported on e-waste are stark. According to the latest *Global E-Waste Monitor*, in 2019, the world produced 53.6 million metric tons of e-waste, which amounts to 7.4 kg per capita (Forti *et al.,* 2021). Of this total, 6.7 million metric tons come from screens and monitor, 4.7 million metric tons from small ICT equipment and 10.8 million metric tons from temperature exchange equipment, such as the cooling devices used on an industrial scale in data centre complexes (Ibid.). Estimates say global rates of e-waste production will rise to over 74 million metric tons per year by 2030 (Murthy and Ramakrishna, 2022). Furthermore, the unfortunate reality is that most countries are unable to adequately deal with growing levels of e-waste, with many turning to illegally exporting e-waste to developing countries (Tidey, 2019).

Recently, there has been a policy push towards promoting more circular production ecosystems (Murthy and Ramakrishna, 2022). Interestingly, this creates opportunities for resource extraction, due to the high levels of precious metals and valuable resources present in ICT e-waste. "Urban mining" of e-waste could significantly ease the environmental impacts associated with developing new devices and augmenting and maintaining devices already in use (Ibid.). E-waste, therefore, represents a challenge to the environmental sustainability of expanding data for learning ecosystems, but also raises the potential for the creation of new industries extracting the resources required to develop and maintain them.

Promoting sustainable consumption patterns to all stakeholders across the ecosystem

How, then, do we incentivize sustainable behaviour to ensure data for learning models support environmental well-being? From the social well-being perspective, it is important that investments in data infrastructure that supports the expansion of data-fuelled learning models – especially those that rely on big data – take the necessary measures to mitigate the environmental impacts of energy-consuming big data analytics. Likewise, the reliance on such big data models must consider and protect those employed to clean big datasets, who may experience sensitive and traumatic material in the cleaning process (Perrigo, 2022). Therefore, the use of locally sourced, smaller datasets should also be considered to support sustainable data for learning ecosystems.

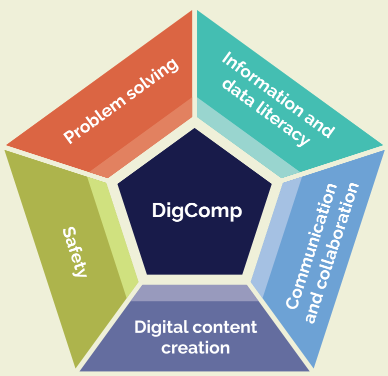
This small over large viewpoint is also creeping into initiatives via a data minimalism mindset, which has, in part, been influenced by the environmental impacts discussed earlier. The European Edtech Alliance highlighted several examples of this mindset from across Europe. From a corporate responsibility angle, some companies have revised their digital strategies in an attempt to discover what level of minimalism can sufficiently maintain their services, under the belief that more data does not automatically lead to improved results (Softtek, 2021). Regarding policy and education strategy, Educa Swiss in Switzerland has been working to tie the general Swiss government data strategy to a better understanding of the competencies required to underpin a data for learning ecosystem, as well as data for research (Educa Swiss, 2019). This includes having national standards for understanding the reason for data use (needs, immediate action requirements, roles and responsibilities, etc.).

It is important to improve capacities in this regard across the data for learning ecosystem rather than simply concentrating on the supply side aspects. All stakeholders should be aware of the physical costs that come with their digital actions. Frameworks such as the European Commission's European Digital Competence Framework for Citizens offer useful insights into the competencies required across the ecosystem to ensure that the environmental challenges highlighted above are not only taken into account, but actively mitigated as efficiently as possible (Vuorikari *et al.,* 2022). In this regard, when and wherever necessary, regular training for relevant stakeholders could facilitate a continuous learning process that addresses the constantly evolving reality associated with shifting commitments, targets, and technological advancements and efficiencies.

Finally, as discussed, e-waste represents another key challenge that should be accounted for, particularly when the introduction of new digital devices into a learning environment is being considered or mandated. Taking a more circular approach could offer significant savings on e-waste generation. Developing new devices with ease of maintenance, upgrades, and ultimately the reclaiming and recycling of used components, resources and materials offers a means for reducing levels of e-waste, as well as potentially laying the foundations for productive secondary industries (eReuse, n.d.; Bel *et al.,* 2019; Rreuse, 2017). This circular vision could also apply to data centre energy usage, with reclaiming and reusing the heat created by servers offering a way to mitigate traditional losses, such as has been implemented by Meta at the Odense Data Center in Denmark (Meta, 2020).

Box 20. DigComp 2.2: The European Digital Competence Framework for Citizens

The European Commission's European Digital Competence Framework for Citizens offers an insight into digital literacy and digital competences. It offers examples of knowledge, skills and attitudes across a broad range of aspects, such as evaluating data, managing your digital identity, and protecting your personal data and privacy, among others. Competence area four of the framework focuses on safety, and includes an awareness of the environmental impact of digital technologies and the impact their use may have (Directorate-General for Education, Youth, Sport and Culture [European Commission], 2019).



Information and data literacy	1.1. Browsing, searching and filtering data, information and digital content 1.2. Evaluating data, information and digital content 1.3. Managing data, information and digital content
Communication and collaboration	2.1. Interacting through digital technologies 2.2 Sharing information and content through digital technologies 2.3. Engaging in citizenship through digital technologies 2.4 Collaborating through digital technologies 2.5. Netiquette 2.6. Managing digital identity
Digital content creation	3.1. Developing digital content 3.2. Integrating and re-elaborating digital content 3.3. Copyright licences 3.4. Programming
Safety	4.1 Protecting devices 4.2 Protecting personal data and privacy 4.3 Protecting health and well-being 4.4 Protecting the environment
Problem solving	5.1. Solving technical problems 5.2. Identifying needs and technological resources 5.3. Creatively using digital technologies 5.4. Identifying digital competence gaps

5

Key findings, recommendations and future developments

Today, education and training systems around the world are struggling to deploy their agency to steer the data revolution on their own terms towards targeting persistent education challenges and strengthening system readiness for a rapidly changing world.

There is a critical need to **establish benchmarks and standards** for utilizing data with appropriate agency. To do this, we must engage parties from across the entire ecosystem – including policy-makers, firms, organizations, education institutions and, most of all, learners, teachers, school staff and others – to build common definitions, practices and a visionary framework.

In 2021, the Data Futures Platform of the United Nations Development Programme (UNDP) published eight data principles, and illustrated their alignment with existing international initiatives and frameworks. These principles are: (1) safeguard personal data; (2) uphold the highest ethical standards; (3) manage data responsibly; (4) make data open by default; (5) plan for reusability and interoperability; (6) empower people to work with data; (7) expand frontiers of data; and (8) be aware of limitations. UNESCO's Right to Education Initiative is building on this existing work by specifically discussing the impact of data use in education on human rights in forthcoming analytic work.

Tying these frameworks to the broader mission of the Broadband Commission, this report underscores that, in the digital age, a high proportion of people are shut out of learning and economic activity due to barriers to digital and data equity. Data for learning should be seen as part of the vision that education is a public good that helps individuals reach their full potential.

We must harness the power of the digital revolution, including the data revolution, to ensure that equitable and quality education and lifelong learning are provided as a human right, with a particular focus on the most marginalized. This has been affirmed by the United Nations Secretary-General's Our Common Agenda (United Nations, 2021), the United Nations Roadmap for Digital Cooperation (United Nations, 2020), the Rewired Global Declaration on Connectivity for Education (UNESCO, 2022*d*) and the International Commission on the Futures of Education (2021), and was at the centre of the Transforming Education Summit discussions.

As this report shows, designing, implementing, governing and monitoring such information systems is no simple task in the complex education sector, of which many asymmetries impede synergized approaches to data for learning.

Asymmetry of information: Swift and sweeping social changes vs slow and siloed data flows

Education systems face growing pressures to rapidly respond to the many transitions unfolding around the world, from demographic and labour-market shifts to numerous digital revolutions and climate change. However, the sector does not always have or use available information to anticipate the impact of these changes. As a result, education management often becomes a reactive rather than proactive process, which renders it unprepared to mitigate crises.

Asymmetry of skills: Big demands for "data-driven" vs little literacy for "data-informed"

There is growing pressure to use data in decision-making at every level of the education system. Therefore, every actor in the education system should have the necessary data-related competences to ensure effectiveness, accountability, compliance, privacy and security in the use of education data to improve quality and equity in all learning settings. While data-*driven* decisions tend to be based on a direct surface reading of data, data-*informed* approaches are inferred from a contextualized and critical interpretation of education data – one that balances the data with human insights and shared principles of ethical and responsible use.

Asymmetry of sovereignty: Local legal frameworks vs global data flows

Data in education are not static. They are collected, processed, transferred, stored, combined, separated, archived and destroyed – often in different settings. These settings may not be the same classroom, school district, country or even continent. A growing concentration of data is in the hands of a few large corporations and countries located in the global North. Without adequate legislation or public awareness, the current imbalance in data power could restrict user autonomy over their data and compromise national data sovereignty and security. To avoid undermining education as a universal human right, it is crucial to assess how their support can benefit resource-scarce education systems.

Therefore, the WGD4L proposes the following recommendations for addressing the asymmetries and ensuring the potential benefits of data in education to be unlocked:

1. Develop and implement a **whole-of-government and whole-of-ecosystem vision and strategy** on the use of data for learning, grounded in a rigorous understanding of the potential opportunities, benefits, limitations, risks and impact.

Data collection should always stem from specific, transparent and targeted uses, such as informing teaching practices, learning environments and strategic management, with an exclusive focus on the data required to implement that specific use. Furthermore, data should not be collected solely because it is technologically feasible or easy to do so, since collecting data without an intentional and transparent intended use can lead to hyper-datafication or surveillance that may limit the autonomy of teachers and learners, reproduce social inequalities or jeopardize individual security. Data collection should directly support the formulation of contextualized observations and recommendations to be used in combination with human insights. Learners are key creators of these insights, and should be empowered to participate in data collection and analysis processes as central players in holistic strategies for data for learning ecosystems.

There is also a need to unify interoperable data processes to bring nationally sourced and partner-sourced data into local analysis. A unified approach will not only help at the local level, but also offer sector-wide support for decision-making at the district and provincial levels. For reasons including poor human resource capacities and the inability of education systems to fully integrate and apply learning data, many government actors do not fully leverage the many data sources available to them when making decisions and defining education policy. We must therefore facilitate multi-stakeholder partnerships to deliver the skills required to take full advantage of the new possibilities learning data can offer. Incentivizing a whole-of-government approach creates an environment conducive to overcoming the obstacles preventing many governments from efficiently using data to support the public and common good for the benefit of learners.

Furthermore, inherent tensions exist between privacy and transparency, and there is a need to ensure that policy is data-*informed* rather than data-*driven* so that decisions are not made upon partial or biased evidence. Following the lead of the UNESCO Recommendation on the Ethics of AI, as well as the Rome Call for AI Ethics of 2022, governance in the context of data for learning should therefore incorporate transparency and accountability as active aspects of protection, while also enabling effective monitoring of impact, enforcement and redress. This will require broad partnerships that can also reach beyond national borders to ensure effective regulation at the national and global levels, with an understanding of the inherent tensions in data ecosystems and the added sensitivity related to learning environments. Any strategy seeking to engage the true value of data for learning should also understand the limits of the data and seek to build contextualized, domain-informed and critical interpretations that balance data with common sense and shared values.

2. Establish a **sustainable financing strategy** for data for learning grounded in national financial resources that centres learners' interests, contributes to the common good, minimizes environmental impact and is sustained by strategic multi-stakeholder partnerships.

A key challenge standing in the way of educational data becoming a truly transformational technology for learning is financing. Collecting and processing educational data can be prohibitively expensive for smaller players such as schools. Duplicating data collection across disparate management systems is a costly and time-wasting process. Adopting a whole-of-government and whole-of-ecosystem approach to fully appreciate the benefits of an inclusive and equitable lifelong learning ecosystem requires a sustainable financing model that covers both hard and soft elements. This will be expensive, with hard elements ranging from hardware and equipment, such as for data storage, in data centres to ongoing monthly and annual energy consumption, hardware cooling and other costs. The soft elements of data ecosystems cover a broader array of considerations, including database management costs, data and digital literacy training, software development and, where necessary, IT staff who can manage and maintain digital systems.

At the moment, many countries simply do not have the connectivity or infrastructure required to benefit from an advanced data for learning ecosystem. In fact, many countries do not even have tangible and exact data on network coverage and connectivity. Even high-income countries face challenges relating to data architecture when seeking to exploit frontier technologies, and continue to struggle to expand the existing digital infrastructure to remote communities. Without significant and continuing investment, implementing inclusive and equitable data for learning ecosystems that transform lifelong education opportunities for all will prove difficult, if not impossible.

Furthermore, it is imperative that all sustainable financing mechanisms provide for the public good and deliver ongoing benefits to learners. The complexity and breadth of the hard and soft elements of effective data for learning ecosystems must come together to build powerful capabilities that can be harnessed for the good of education, rather than be seen as proprietary concerns. Therefore, underpinning the costs and complexities involved in financing data for learning ecosystems are the risks attached to bringing them into already under-resourced education ecosystems. This means that sustainably financing these new and evolving ecosystems is not just about provision, but also about the adequate mitigation of the challenges and risks discussed throughout this report.

Finally, it is also important to consider that many of the hard and soft elements of digital transformation also come with costly environmental impacts. As such, environmental sustainability must be built into all long-term data for learning initiatives. Citizen engagement and learner empowerment are key components of building sustainability into all data for learning plans and strategies, as youth can contribute to data availability and insights that can enhance sustainable educational systems. Students should be empowered to collect real-time data themselves and use them to promote sustainability in their schools and communities based on sound data and data analytics.

3. Strengthen critical **data literacy and skills** at all levels of the education ecosystem to facilitate improved regulation and inclusive innovation through effective implementation and monitoring of education data policies and practices.

Data literacy and skills are required across the education ecosystem, from classroom learners to government ministers. Insufficient digital skills and literacy form a tangible pillar of the digital divide, preventing accessible digital transformation on a variety of scales. Many governments are currently unable to fully exploit the transformational potential that data for learning offers as they do not have sufficiently skilled human resources to do so. Many teachers remain disconnected from professional development pathways new data-*informed* education practices are opening up, and many learners are unable to take advantage of the new digital education tools now available.

For data learning ecosystems to be fully inclusive, efforts should be made to strengthen data literacy skills, including those of learners and students; teachers, educators, instructors and assistants; and administrators, governors, school leadership teams and policy-makers. Stakeholder support could follow the UNESCO framework, focusing on improving competencies including data and media literacy, as well as "… the ability to access, manage, understand, integrate, communicate, evaluate and create information safely and appropriately through digital technologies for employment, decent jobs and entrepreneurship" (Law *et al.,* 2018). Such efforts build upon the work of the Broadband Commission Working Group on Smartphone Access, which called for a shift of focus from the "coverage gap" to the "usage gap" in September 2022 to reorient efforts away from access alone and towards meaningful connectivity through ecosystem literacy.

Effective data literacy across the ecosystem from top to bottom will help facilitate improved regulation through more effective participation and monitoring of data for learning policies and practices. Importantly, this includes empowering learners to collect data locally to facilitate autonomous pathways towards improving their understanding of local and global phenomena through the data they are able to collect and analyse themselves.

This will not, however, be a static endeavour. As data as a technology and tool for education continues to evolve, so will the skills required to understand and engage with data across all levels of the education ecosystem. The capacity-building required to break down one of the pervasive pillars of the digital divide should therefore also address the need to understand an evolving landscape and the skills required to traverse it, and support infrastructure that can develop, assess and certify those skills. This will require the development of active assessment, measurement, monitoring and evaluation frameworks to ensure the continued and effective development of digital skills and literacy, even as the skills themselves evolve.

Beyond simply developing data collection, analysis, access and use skills, strengthening data literacy also sits at the heart of other key tensions, such as data privacy, ownership, consent, transparency, accountability and security.

4. Prioritize the potential benefits of data to **transform education by targeting education's enduring obstacles** to assist informed and inclusive quality learning, teaching, administration, management, planning and financing.

This report has identified key challenges and risks associated with data for learning practices, not least the simultaneous existence of more and bigger data than have ever been collected before, and the unshakeable reality that educational data will only ever be able to capture certain aspects of educational practices. This dichotomy pulls at the potential effectiveness of inclusive and equitable data for learning ecosystems by shrouding information in data, and at the same time offering key insights from analysis. In this environment, it is easy to lose sight of the real objectives of an endeavour, seeing only what the data shows us. It is important to remember that data is not information, and that information is not wisdom.

A data for learning ecosystem that delivers inclusive and equitable lifelong learning for all will work hand in hand with human understanding to ensure that data-*informed* policies govern with data rather than being governed by data. Consequently, it is important to focus on the benefits of data-*informed* learning for global efforts to transform education and tackle some of its most persistent problems. In this regard, and to build on the digital literacy outlined in the previous recommendation, policy-makers also need to understand the potential benefits data for learning policies offer, and the possible and plausible harms. This includes a strong, functional understanding of the legal frameworks in which data for learning policies will be considered, so as not to overlook unintended consequences and to fully comprehend associated direct and indirect costs. This will require collaboration between policy-makers and data science, education and privacy experts to support informed decision-making processes and effective data for learning policy-making.

Care must also be taken to ensure that the data itself is both inclusive and representative. Data governance models that support efforts to transform global education will require policies to promote and increase diversity and inclusiveness within datasets to protect against any bias or values that may exist in the contexts where the data were collected and produced. Furthermore, mechanisms for disclosing and combating any cultural, economic or social prejudices present in data, either by design or negligence, are vital, particularly in areas where data are scarce.

5. Harness **multilateralism, solidarity** and **international cooperation** to bridge the digital divide, nurture local data capacities and promote open authentic data to support the development and implementation of better tools for the education sector, including international standards and norms over education data regulation, literacy, cybersecurity, governance and ethics.

As this report has explored, addressing the challenges associated with inclusive data education ecosystems to fully reap the potential benefits of data for learning will require significant investment. However, the barrier to entry for countries with existing strong digital infrastructure and high levels of skilled human capital is much lower than for lower-resourced countries. Countries without strong digital infrastructure see contexts where resources are repeatedly consumed by collecting multiple instances of the same learning data that often exist in silos across education management systems.

Echoing this unfortunate reality are increasing claims of national data sovereignty that seek to push back on the open data movement and close cross-border data flows. International solidarity and cooperation are vital if learners' human rights are not to be infringed in the pursuit of maintaining national sovereignty over datasets. This is particularly true in the cases where data sovereignty is being pursued as a defence against data colonialism, which sees companies and institutions in the global North taking ownership over data from the global South. There is a critical balance to be struck between regulating data ownership to prevent the exploitation or commercialization of learner data, particularly by distant companies that have no positive local economic impact, and incentivizing data openness to democratize data as an educational resource and improve transparency in education governance.

Open-licensed data, a legal condition that guarantees permissions for sharing, transforming and combining data in common and machine-readable formats, is a critical area that can maximize the flexibility of public use of and engagement with data, and unleash many uses that promote data literacy and skills building in an equitable way. Open data enables individuals and organizations to access and reuse data to innovate and collaborate in a transparent context, which could promote citizens and governments to work in cooperation to plan and monitor improved public services and local businesses to develop context-relevant digital tools.

Future development

The work of this two-year Broadband Commission WGD4L and this report have taken important steps towards defining and refining the meaning of data for learning. This includes proposing recommendations on promoting inclusive, equitable and successful data for learning ecosystems. However, with data generation and cross-border data flows rising globally and digital data-*driven* practices increasingly influencing education ecosystems, more work needs to be done to share knowledge of actionable measures that governments can take to respond to the recommendations of this report.

The most challenging issue for education systems will be to figure out the contours of data for learning and data management systems that allow for a balance between human insights and technology-enabled emerging practices. There is a need for contextualized solutions that are forward-looking while remaining sensitive to capacity constraints and local contexts. In low-resource and fragile contexts, investments in developing baseline digital skills and data literacies may be more effective than trying to adopt advanced EMIS architectures. Moreover, understanding the evolving dimensions of the right to education in an increasingly digital learning landscape deserves nuanced discussion.

Generative AI

Generative AI is a category of AI algorithms that generate seemingly new and realistic content based on the data they have been trained on. The content can be in image, audio or text form, meaning it can be used to generate everything from portraits, landscape paintings, music and voice-overs to poetry, screenplays, translations and emails. Generative AI requires a system and set of algorithms to operate. In the case of sophisticated chatbots, such as ChatGPT, this system is a large quantity of text data. Using mathematical probabilities, a large language model (LLM) is then trained on this large dataset to read, recognize, summarize, translate, predict and ultimately, generate new text.

From machine translation to virtual assistants, adaptive tutoring and assessment systems, generative AI could see many promising use cases in education, especially when studies show that up to 70 per cent of children in poorer countries are unable to read a basic text by age 10. However, the use of generative AI in education is also raising fears that these tools will increase cyber incidents and data breaches, replace teachers, encourage plagiarism, or amplify existing biases and discrimination in education systems through biased training datasets. These tools are producing massive amounts of new Internet data, upon which the next model will be trained, which marginalizes the voices of the already marginalized.

Though they may not have human minds, humans are behind every aspect of these tools, from creation to implementation. From the creators who write the algorithms and decide what to include in training datasets to the design of the tool and intentions behind its use, individuals – and their context, language, gender and experiences – feed into the development of generative AI tools. Therefore, it is critical to require transparency from the individuals and companies creating these tools to avoid bias in datasets, explain decision-making, and build trust around the intended applications and impact of these tools.

LLMs have high computational demands, which can result in high energy consumption. Hence, energy-efficient hardware and shared (e.g. cloud) infrastructure based on renewable energy are crucial for their environmentally sustainable operation and scaling needed in the context of education. For model training and updates, only data that has been collected and annotated in a regulatory compliant and ethical way should be considered. Therefore, governance frameworks that include policies, procedures and controls to ensure appropriate use of such models are key to their successful adoption. Likewise, for the long-term trustworthy and responsible use of the models, transparency, bias mitigation and ongoing monitoring are indispensable.

AI for monitoring and optimizing education systems

The potential of AI in monitoring and optimizing education policy and systems is immense, ushering in a new era of data-driven decision-making, personalized interventions and transparent policy formulation. As highlighted in the case studies from India and Australia (Box 20), AI's predictive analytics and sophisticated algorithms enable early identification of at-risk students, leading to targeted interventions and improved retention rates. Additionally, AI-driven insights empower educational policy-makers to craft evidence-based strategies that address challenges ranging from student dropout rates to resource allocation optimization. However, while these advancements underscore the transformative capabilities of AI, it is imperative to recognize that more work needs to be done. Sharing actionable measures and fostering collaboration among governments, institutions and AI experts are essential to maximize the potential of AI in education. Governments can respond to the recommendations of such initiatives by embracing technology, enhancing digital skills and ensuring ethical AI practices. The following case studies exemplify the ongoing efforts in harnessing AI's potential for informed policy decisions and optimized educational outcomes.

Box 21. The transformative potential of AI in monitoring education policy and systems

AI presents a revolutionary potential for monitoring and managing education policy and systems, offering diverse applications that redefine educational system optimization and monitoring. Through predictive analytics, resource optimization and personalized interventions, AI fosters an environment of proactive decision-making, transparency and data-driven policy formulation while keeping "human-in-the-loop". This transformative potential is exemplified by case studies from India and Australia, where AI-driven initiatives are showing real promise in rethinking the landscape of educational monitoring systems.

Case Study: Enhancing student retention in Andhra Pradesh, India (NITI Aayog, 2018)

Collaborating with Microsoft, the Andhra Pradesh Government has embarked on an innovative journey to combat the challenge of school dropouts. By leveraging Azure Machine Learning, a cutting-edge platform, the state is delving beyond conventional data analysis. The system deciphers intricate variables such as gender, socio-economic demographics, academic performance, school infrastructure and educator proficiency. These components converge to unveil hidden patterns contributing to student attrition.

The true strength of this initiative lies in its actionable insights. Armed with a profound understanding of dropout determinants, education officials can enact targeted interventions. Through a multifaceted approach of programmes and counselling, encompassing both students and parents, the Andhra Pradesh Government fosters a holistic support network. The results are impressive. The initiative effectively identified around 19,500 potential dropouts from government schools in the Visakhapatnam district for the academic year 2018/19. This data-driven approach exemplifies AI's capacity to address complex societal challenges, paving the path for strategic policy interventions.

Case Study: Empowering Governance through AI in New South Wales, Australia (Digital.NSW, n.d.)

The New South Wales (NSW) Government's AI strategy serves as a case study of progress in governance and service delivery. Embracing AI's potential, the strategy seeks to automate processes, amplify efficiency and allocate resources judiciously. AI functions as an invaluable decision support tool, aiding rather than replacing human judgment. Ethical considerations remain paramount, and any AI-informed decision is subject to swift and efficient review, ensuring accountability and citizen-centric governance.

The strategy's development has been fortified by extensive consultations, culminating in five key themes: building public trust, digital upskilling, enhancing data capabilities, innovating procurement processes and fostering collaboration. The strategy's iterative approach has led to tangible accomplishments, including the creation of an AI Assurance Framework, the establishment of Australia's first AI Advisory Committee and the publication of AI case studies. These actions reflect the NSW Government's commitment to ethical, effective and transparent AI utilization.

As the NSW Government continues its journey of AI integration, a strong focus remains on awareness, formalizing assurance frameworks, clarifying oversight mechanisms and nurturing sustained public engagement. This enduring commitment showcases AI's potential to enhance governance and amplify service delivery, while keeping citizens' well-being at the forefront.

By harnessing the power of AI, both case studies from Andhra Pradesh and NSW are illuminating the transformative potential that AI holds for monitoring and optimizing education policies and systems, redefining educational paradigms, and paving the way for a more equitable and empowered future.

Synthetic data

Engineers generate synthetic data based on a smaller sample of real data that is labelled with all the aspects deemed relevant for the AI models to train on, and a set of rules that seek to counteract any obvious and known biases in the original dataset. However, taking too optimistic a view of synthetic data for creating data ecosystems, without bias or privacy issues, overlooks the reality that it is still individuals who are making decisions about what data to include and exclude, and how to analyse the data, with those choices based on what is deemed important or relevant by those individuals. If people are making decisions on which of these datasets should be built, which problems they should solve and what real-world data they should be based on, we will never be able to fully remove bias. Furthermore, as synthetic data are based on smaller samples, such data may not only reproduce the patterns and biases drawn from the data, but amplify them too.

In a worst-case scenario, we could get an echo chamber effect, whereby AI feeds the AI, and models that develop and control key aspects of our world – the information we consume, the digital worlds we frequent, the recommended learning paths and learning products we receive – increasingly respond to an internal logic divorced from the reality we inhabit. If the synthetic dataset is not grounded in (or perhaps made from) a rigorous understanding of the most recent underlying human phenomenon, such as the differences between what people say and do, or the unexpected influence of tangential variables in our lives on the actions we take, it risks simulating a social world that short-changes reality in ways that could cause real harm to individuals. And this is before we even begin to contemplate more nefarious uses of synthetic data, such as deepfakes or misinformation on a massive scale.

Synthetic data also raise complicated issues relating to privacy and consent. From the legal perspective, as synthetic data are often not data relating to a natural person (under the GDPR), or to a particular consumer or household (under the California Consumer Privacy Act [CCPA]), or to an individual (under the Health Insurance Portability and Accountability Act [HIPAA]), then they are not considered to be personally identifiable or sensitive information. Such data are therefore outside the scope of these privacy laws. As such, synthetic data are not held to the same accountability principles and profit from being detached from "natural" individuals or legal structures. In the absence of clear associations, synthetic data rights should be rethought and protected by design and regulations, especially those concerning consent, privacy, accountability and explicability. Otherwise, the unethical adoption of synthetic data in education could go unchecked before scaling becomes irreversible.

Data for machine learning and unlearning

It is straightforward to delete a customer's data from a database and stop using it to train future models. But what about models that have already been trained using an individual's data? These are not necessarily safe; it is known that individual training data can be exfiltrated from models trained in standard ways via model inversion attacks (Veale *et al.,* 2018). Regulators are still grappling with when a trained AI model should be considered to contain individuals' personal data in the training set, and what the potential legal implications may be.

Data protection and privacy have been the subject of much discussion as more and more individuals come to realize just how much personal information they are sharing through the countless apps and websites they regularly visit. Many people are concerned. Recent government initiatives such as the European Union's GDPR are designed to protect individuals' data privacy, with a core concept being "the right to be forgotten". The bad news is that it is generally difficult to revoke things that have already been shared online or to properly delete such data. Facebook, for example, launched an "off-Facebook activity" tool (previously called "Clear History"), which the company says enables users to delete data that third-party apps and websites have shared with Facebook. But, as the MIT Technology Review notes, "it's a bit misleading – Facebook isn't deleting any data from third parties, it's just de-linking it from its own data on you." Machine learning is increasingly viewed as exacerbating this privacy problem (Synced, 2020). Machine learning applications are driven by data, and this can include collecting and analysing information, such as personal emails or even medical records. Once fed into a machine learning model, such data can be retained forever, putting users at risk of all sorts of privacy breaches.

Switching to a researcher's perspective, a concern is that if and when a data point is actually removed from a machine learning training set, it may be necessary to retrain downstream models from scratch. In a new paper, researchers from the University of Toronto, the Vector Institute and the University of Wisconsin–Madison propose SISA training (Bourtoule *et al.,* 2020), a new framework that helps models "unlearn" information by reducing the number of updates that need to be computed when data points are removed. "The unprecedented scale at which machine learning is being applied on personal data motivates us to examine how this right to be forgotten can be efficiently implemented for machine learning systems", the researchers explain. Having a model forget certain knowledge requires that some particular training points be made to have zero contribution to the model. But data points are often interdependent and can hardly be removed independently. Existing data also work continuously with newly added data to refine models. One solution is to understand how individual training points contribute to model parameter updates. But as previous studies have shown, this approach is only practical when the learning algorithm queries data in an order that has been decided prior to the start of learning. If a dataset is queried adaptively, meaning that a given query depends on any queries made in the past, this approach becomes exponentially more challenging and thus can hardly scale to complex models such as deep neural networks. All this is to say that every new data-*driven* breakthrough could challenge previously assured safeguards. This is even more important and relevant in the case of lifelong learning, which often involves data relating to minors.

Neurotechnologies and "neurodata"

Recent years have seen a proliferation of global investment in the development of neurotechnologies designed to read and write to the brain (Digital Future Society, 2023 [forthcoming]). While these technologies have primarily been developed for medical applications with evolving applications bringing smaller and increasingly safer devices, neurotechnologies are seeing more and more use in non-medical contexts. For example, being used for augmentation of a brain and nervous system function, rather than a replacement of a lost ability or lost abilities. These new non-medical applications of neurotechnologies reach across sectors including education, meaning "neurodata" could become a governance issue moving forward.

Currently, the neurotechnology landscape is in its infancy, with disparate guidance and regulations, such as from the OECD (2019) and the UNESCO International Bioethics Committee (IBC) (2021), offering broad definitions of both the technologies underpinning the applications and even the types of data they produce. For example, Digital Future Society (2023) notes that there are differences to consider between neurodata, which has been described as first-order data from brain cells or neurons, such as their anatomical components; neural data (Future Privacy Forum, 2021), which IBC (2021) has described as: that recorded directly or indirectly from an individual's brain; mental data or a conglomeration of representations and attitudes corresponding to the experience of thinking, feeling, remembering, etc. (Ienca and Maglieri, 2022); and neuroscience data, which comprise both derived data and metadata to explain the processing steps and analysis used to produce the data (Eke *et al.,* 2022).

While there is an argument that these novel types of data could be incorporated into current and future data governance frameworks, there are advocates claiming that the added sensitivity of these data warrants special attention. There have even been calls for the development of new "neurorights" (Rommelfanger *et al.*, 2022), with Chile already becoming the first country to pass a neurorights bill, which touches on aspects of mental privacy, personal identity, protection from bias and free will, all of which intersect with ethical aspects already intertwined with digital transformation and the increasing use of digital data in education (Hormazábal, 2022).

In the classroom, the use of neurotechnologies have been rare, but there have been trials that again intersect with common challenges underpinned throughout this report. For example, in China, in 2019, parents were concerned with the use of neurotechnology headsets provided by a US and Chinese partnered company called BrainCo, which were designed to help children concentrate and learn (Standaert, 2019). Parents were worried that the devices could be used to violate the privacy of their children and even control them. Furthermore, despite assurances from the company that the data derived from the devices would be stored locally, teachers reported that it was shared with the BrainCo server, raising core issues of transparency and trust. It is worth noting, however, that these concerns surrounding a rapidly developing new technology come in the face of shifting attitudes towards the use of neurotechnologies. Surveys suggest that young people are generally more open to the use of neurotechnologies for enhancement purposes (Sattler and Pietralla, 2022).

It is important, therefore, to consider that with increasing investment driving newer implementations of neurotechnology and shifting attitudes towards their use, neurodata could become a unique governance issue in education data ecosystems of the future.

Areas for further analysis, data collection and policy dialogue

The challenges outlined above are important and represent areas where exploring tensions is likely to be fruitful for data ethics. Going forward, further similar areas can and should be identified. As well as building an in-depth understanding of the interplay throughout and across the education data ecosystem, critical questions still need to be answered if data for learning is to truly contribute to the transformation of education for the common good. These questions include:

→ Where data are being used to serve a particular goal or value, or for social benefit in general, what risks to other values are introduced?

→ Where might uses of data-driven systems that benefit one group or the whole population have negative consequences for a specific subgroup? How do we balance the interests of different groups?

→ Where might applications of data-driven systems that are beneficial in the near term introduce risks in the long term? How do we balance short- and long-term impacts on society?

→ Where might future developments in data-driven systems, including AI, either enhance or threaten important values, depending on the direction they take? (Whittlestone *et al.,* 2019)

Finally, digital divides exist as many people remain removed from the digital transformation of education for a variety of reasons, including connectivity, access to devices and skills gaps. However, once these digital divides are bridged, a data divide emerges. Due to inequalities in capacity, different parts of the world are at very different stages of developing the necessary safeguards and protections to ensure that learner data are private, secure, and protected from unethical commercialization or compromise. This data divide is drawn across socio-economic lines that leave learners in lower-income areas vulnerable to rights abuses and even ignorant of the knowledge that they possess data rights.

It is clear, therefore, that the data divide is a rights and equity issue, as well as an issue of what is perceived as valuable in the education sector. Even if data are reliable, secured and trustworthy, they cannot capture a full picture of learning. If data for learning is to truly help transform education for the common good, they must be socially contextualized, used skilfully, safe and secure, and, above all, they must serve the primary purpose of improving teaching and learning experiences.

References

Allegretti, A. 2020. *A-Level results: Government accused of 'baking in' inequality with 'boost for private schools*. Sky News. https://news.sky.com/story/35-of-a-level-results-downgraded-by-one-grade-figures-reveal-12048251 (Accessed 14 August 2023.)

Ang, C. 2020. *What does 1 GB of mobile data cost in every country?* Visual Capitalist. https://www.visualcapitalist.com/cost-of-mobile-data-worldwide/ (Accessed 14 August 2023.)

Antoninis, M. 2018. *A global framework to measure digital literacy*. UNESCO Institute for Statistics. https://uis.unesco.org/en/blog/global-framework-measure-digital-literacy (Accessed 17 August 2023.)

Arias Ortiz, E., Eusebio, J., Pérez Alfaro, M., Vásquez, M. and Zoido, P. 2021. *Education Management and Information Systems (SIGEDs) in Latin America and the Caribbean: The Road to the Digital Transformation of Education Management*. Washington, DC, Inter-American Development Bank. https://publications.iadb.org/publications/english/viewer/Education-Management-and-Information-Systems-SIGEDs-in-Latin-America-and-the-Caribbean-The-Road-to-the-Digital-Transformation-of-Education-Management.pdf

Arias Ortiz, E., Giambrun, C., Muñoz Stuardo, G. and Pérez Alfaro, M. 2021. *The Road to Educational Inclusion: Four Steps to Develop Systems to Protect Educational Pathways*. Washington, DC, Inter-American Development Bank. https://publications.iadb.org/publications/english/viewer/The-Road-to-Educational-Inclusion-Four-Steps-to-Develop-Systems-to-Protect-Educational-Pathways.-Step-2-Designing-Early-Warning-Systems.pdf

Arushayan, Y. 2016. *Environmental Impacts of ICT: Present and Future*. Stockholm, KTH Royal Institute of Technology. https://www.diva-portal.org/smash/get/diva2:933594/FULLTEXT01.pdf

Behzadi, Y. 2021. *Synthetic data to play a real role in enabling ADAS and autonomy*. Automotive World. https://www.automotiveworld.com/articles/synthetic-data-to-play-a-real-role-in-enabling-adas-and-autonomy/ (Accessed 14 August 2023.)

Bel, G., van Brunschot, C., Eaesen, N., Gray, V., Kuehr, R., Milios, A., Mylvakanam, I., Pennington, J. and The Secretariat of the Basel and Stockholm Conventions. 2019. *A New Circular Vision for Electronics: Time for a Global Reboot*. Cologny/Geneva, Switzerland, World Economic Forum. https://www3.weforum.org/docs/WEF_A_New_Circular_Vision_for_Electronics.pdf

Belkhir, L. and Elmeligi. A. 2018. Assessing ICT global emissions footprint: Trends to 2040 & recommendations. *Journal of Cleaner Production*, Vol. 177, pp. 448–463. https://www.sciencedirect.com/science/article/abs/pii/S095965261733233X

Berridge, E. R. 2020. *Impact of Covid-19 on summer exams: Statement made by Baroness Berridge on 23 March 2020*. UK Parliament. https://questions-statements.parliament.uk/written-statements/detail/2020-03-23/HLWS170 (Accessed 14 August 2023.)

Bertino, E. and Lim, H.-S. 2010. Assuring data trustworthiness – concepts and research challenges. W. Jonker and M. Petković (eds.), *Secure Data Management*. Heidelberg, Springer Berlin, pp. 1–12.

Blackboard. 2022. *Achieving Its "Student First" Mission Using Blackboard*. Keiser University. https://www.blackboard.com/sites/default/files/2022-10/Keiser_University_Success_Story.pdf

Bourtoule, L., Chandrasekaran, V., Choquette-Choo, C. A., Jia, H., Travers, A., Zhang, B., Lie, D. and Papernot, N. 2020. Machine unlearning. *IEEE Symposium of Security and Privacy*, Vol. 42. http://arxiv.org/abs/1912.03817

Brazil, General Personal Data Protection Law, Law No. 13.709 of 14 August 2018 (as amended by Law No. 13.853 of July 2019).

Brazilian Internet Steering Committee. 2022. *ICT in Education Survey 2021*. São Paulo, Grappa Marketing Editorial. https://cetic.br/media/docs/publicacoes/2/20221121124807/executive_summary_ict_in_education_2021.pdf

Broadband Commission for Sustainable Development (Broadband Commission). 2011. *Working Group on Multilingualism*. https://broadbandcommission.org/working-groups/multilingualism-2011/ (Accessed 14 August 2023.)

Broadband Commission. 2017. *Working Group on Education, 2017*. https://broadbandcommission.org/working-groups/education/ (Accessed 14 August 2023.)

Broadband Commission. 2019. *Working Group on Child Online Safety*. https://broadbandcommission.org/working-groups/child-safety-online-2019/ (Accessed 14 August 2023.)

Broadband Commission. 2020. *Working Group on School Connectivity*. https://broadbandcommission.org/working-groups/school-connectivity-2020/ (Accessed 14 August 2023.)

Broadband Commission. 2021*a*. *Connecting Learning Spaces: Possibilities for Hybrid Learning*. Geneva, International Telecommunication Union (ITU). https://broadbandcommission.org/wp-content/uploads/dlm_uploads/2021/09/Digital-Learning-Report-Broadband-Commission.pdf

Broadband Commission. 2021*b*. *Working Group on Digital Learning*. https://broadbandcommission.org/working-groups/digital-learning-2021/ (Accessed 14 August 2023.)

Broadband Commission. 2022*a*. *Working Group on AI Capacity Building*. https://broadbandcommission.org/ai-capacity-building/ (Accessed 14 August 2023.)

Broadband Commission. 2022*b*. *Working Group on Data for Learning*. https://broadbandcommission.org/data-for-learning/ (Accessed 14 August 2023.)

Broadband Commission. n.d. *Achieving the 2025 Advocacy Targets*. https://www.broadbandcommission.org/advocacy-targets/ (Accessed 14 August 2023.)

Bruns, B., Evans, D. and Luque, J. 2011. *Achieving World-Class Education in Brazil: The Next Agenda*. Washington, DC, World Bank. https://doi.org/10.1596/978-0-8213-8854-9

Carroll, A., Forrest, K., Sanders-O'Connor, E., Flynn, L., Bower, J.M., Fynes-Clinton, S., York, A. and Ziaei, M. 2022. Teacher stress and burnout in Australia: Examining the role of intrapersonal and environmental factors. *Social Psychology of Education*, Vol. 25, pp. 441–469.

Castle, S. 2020. *Boris Johnson retreats in a U.K. exam debacle*. The New York Times. https://www.nytimes.com/2020/08/17/world/europe/england-college-exam-johnson.html (Accessed 14 August 2023.)

Chaudry, R. and Tajwar, A. W. 2021. The Punjab Schools Reform Roadmap: A medium-term evaluation. F. M. Reimers (ed.), *Implementing Deeper Learning and 21st Century Education Reforms: Building an Education Renaissance After a Global Pandemic*. Cham, Switzerland, Springer, pp. 109–128.

Check Point Research. 2022. *Cyber Attack Trends: Check Point's 2022 Mid-Year Report*. Tel Aviv, Check Point Software Technologies Ltd. https://resources.checkpoint.com/cyber-security-resources/cyber-attack-trends-report-mid-year-2022

Council of Europe. 2023. *European Qualifications Passport for Refugees*. Council of Europe Portal. https://www.coe.int/en/web/education/recognition-of-refugees-qualifications (Accessed 14 August 2023.)

Coursera. 2022. *Global Skills Report 2022*. https://www.coursera.org/skills-reports/global (Accessed 14 August 2023.)

Daliri-Ngametua, R., Creagh, S. and Hardy, I. 2021. Data, performativity and the erosion of trust in teachers. *Cambridge Journal of Education*, Vol. 52, No.11, pp. 1–17. https://www.tandfonline.com/doi/abs/10.1080/0305764X.2021.2002811?journalCode=ccje20

Day, E. 2021*a*. *Governance of Data for Children's Learning in UK State Schools: Digital Futures Commission June 2021*. London, Digital Futures Commission, 5Rights Foundation. https://digitalfuturescommission.org.uk/wp-content/uploads/2021/06/Governance-of-data-for-children-learning.pdf

Day, E. 2021*b*. *The education data governance vacuum: Why it matters and what to do about it*. Digital Futures Commission, 5Rights Foundation. https://digitalfuturescommission.org.uk/blog/the-education-data-governance-vacuum-why-it-matters-and-what-to-do-about-it/ (Accessed 14 August 2023.)

Devaux, E. 2021. *List of synthetic data startups and companies – 2021*. Medium. https://elise-deux.medium.com/the-list-of-synthetic-data-companies-2021-5aa246265b42 (Accessed 14 August 2023.)

Digital Future Society. 2023 (Forthcoming). *Humanistic Neurotechnology Development*.

Digital Futures Commission. 2022. *New report finds digital classrooms flout data protection law to exploit children's data for commercial gain*. Digital Futures Commission, 5Rights Foundation. https://digitalfuturescommission.org.uk/blog/new-report-finds-digital-classrooms-flout-data-protection-law-to-exploit-childrens-data-for-commercial-gain/ (Accessed 14 August 2023.)

Digital Futures Commission. 2023. *A Blueprint for Education Data: Realising Children's Best Interests in Digitised Education*. https://digitalfuturescommission.org.uk/wp-content/uploads/2023/03/A-Blueprint-for-Education-Data-FINAL-Online.pdf

Digital Nation Staff. 2022. *Cross-border e-commerce transactions to exceed $2 trillion by 2023: Juniper*. Digital Nation. https://www.digitalnationaus.com.au/news/cross-border-e-commerce-transactions-to-exceed-2-trillion-by-2023-juniper-578349 (Accessed 14 August 2023.)

Digital.NSW. n.d. *Strategy overview*. A NSW Government website. https://www.digital.nsw.gov.au/policy/artificial-intelligence/artificial-intelligence-strategy/strategy-overview (Accessed 21 August 2023.)

Directorate-General for Education, Youth, Sport and Culture (European Commission). 2019. *Key Competences for Lifelong Learning*. Luxembourg, Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/297a33c8-a1f3-11e9-9d01-01aa75ed71a1/language-en

Domo. 2017. *Data Never Sleeps 5.0*. https://www.domo.com/learn/infographic/data-never-sleeps-5 (Accessed 14 August 2023.)

Du Boulay, B., Poulovassilis, A., Holmes, W. and Mavrikis, M. 2018. What does the research say about how artificial intelligence and big data can close the achievement gap? R. Luckin (ed.), *Enhancing Learning and Teaching with Technology*. London, Institute of Education Press, pp. 256–285. https://oro.open.ac.uk/53020/

Duarte, F. 2023. *Amount of data created daily (2023)*. Exploding Topics. https://explodingtopics.com/blog/data-generated-per-day (Accessed 14 August 2023.)

DXtera Institute. 2023. https://dxtera.org/ (Accessed 14 August 2023.)

Edelman, L. 2020. *Facebook's hyperscale data center warms Odense*. Meta. https://tech.facebook.com/engineering/2020/7/odense-data-center-2/ (Accessed 14 August 2023.)

Edelman. n.d. *2021 Edelman Trust Barometer.* https://www.edelman.com/trust/2021-trust-barometer (Accessed 14 August 2023.)

EdSAFE AI Alliance. n.d. https://www.edsafeai.org/ (Accessed 14 August 2023.)

Educa Swiss. 2019. *Data in Education – Data for Education: Principles and Approaches Towards the Development of a Data Use Policy for the Swiss Education Area*. Bern, educa.ch. https://www.educa.ch/sites/default/files/2020-11/data-in-education-management-summary.pdf

Eke, D. O., Bernard, A., Bjaalie, J. G., Chavarriaga, R., Hanakawa, T., Hannan, A. J., Hill, S. L., Martone, M. E., McMahon, A., Ruebel, O., Crook, S., Thiels, E. and Pestilli, F. 2022. International data governance for neuroscience. *Neuron*, Vol. 110, No. 4, pp. 600–612. https://www.sciencedirect.com/science/article/pii/S0896627321009557

EReuse. n.d. https://www.ereuse.org/ (Accessed 14 August 2023.)

FEED Staff. 2020. *Lauren Klien: "There's no such thing as raw data"*. Feed. https://feedmagazine.tv/interviews/lauren-klein-theres-no-such-thing-as-raw-data/ (Accessed 14 August 2023.)

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R. and Warschauer, M. 2020. Mining big data in education: Affordances and challenges. *Review of Research in Education*, Vol. 44, No. 1, pp. 130–160. https://journals.sagepub.com/doi/full/10.3102/0091732X20903304

Forti, V., Baldé, C. P., Ruediger, K. and Bel, G. 2020. *The Global E-Waste Monitor 2020: Quantities, Flows, and the Circular Economy Potential*. Bonn, Geneva and Rotterdam, United Nations University/United Nations Institute for Training and Research, ITU and International Solid Waste Association. https://www.itu.int/en/ITU-D/Environment/Documents/Toolbox/GEM_2020_def.pdf

Garcia, M. 2012. *Higher Education Brasil: History, Policies and a Case Study*. https://www.salzburgglobal.org/fileadmin/user_upload/Documents/2010-2019/2012/495/Session_Document_HigherEducationBrazil_495.pdf

GEM Report. 2022. *Let teachers teach: The dangers of expanding teacher workloads*. UNESCO World Education Blog. https://world-education-blog.org/2017/10/05/let-teachers-teach-the-dangers-of-expanding-teacher-workloads/ (Accessed 14 August 2023.)

Global Partnership for Education. 2021. *Joint Education Sector Monitoring in the Context of COVID-19*. https://www.globalpartnership.org/content/joint-sector-monitoring-context-covid-19-pandemic

Government of Spain. 2019. *Open data to train professionals*. datos.gob.es. https://datos.gob.es/en/blog/open-data-train-future-professionals (Accessed 17 August 2023.)

Greenberg, J., Ringrose, K., Berger, S., VanDodick, J., Rossi, F. and New, J. 2021. *Privacy and the Connected Mind: Understanding the Data Flows and Privacy Risks of Brain-Computer Interfaces*. Washington, DC, The Future of Privacy Forum and IBM. https://fpf.org/wp-content/uploads/2021/11/FPF-BCI-Report-Final.pdf

Grizzle, A. and Singh, J. n.d. *Five Laws of Media and Information Literacy (MIL)*. UNESCO. http://en.unesco.kz/_images/3389.png (Accessed 14 August 2023.)

Group of 20 (G20) Education Working Group. 2022. *G20 Indonesian Presidency Education Working Group Report: Recover Together, Recover Stronger Through Education*. https://unesdoc.unesco.org/ark:/48223/pf0000383466

G20 Education Working Group. 2023 (Forthcoming). *G20 Working Group Report*.

Grupo Interinstitucional de Connectividade na Educação (GICE). 2021. *Guia Conectividade na Educação: Passo a passo para a conectividade das escolas públicas brasileiras*. São Paulo, GICE. https://nic.br/media/docs/publicacoes/13/20210916130704/guia-conectividade-na-educacao.pdf

Gulson, K. N., Sellar, S. and Taylor Webb, P. 2022. *Algorithms of Education: How Datafication and Artificial Intelligence Shape Policy*. Minneapolis, United States, University of Minnesota Press. https://www.upress.umn.edu/book-division/books/algorithms-of-education

Hann, T. 2021. *Synthetic data enables insurers to get more value from AI*. Property Casualty 360°. https://www.propertycasualty360.com/2021/12/24/synthetic-data-enables-insurers-to-get-more-value-from-ai (Accessed 14 August 2023.)

Hardy, I. 2021. The quandary of quantification: Data, numbers and teachers' learning. *Journal of Education Policy*, Vol. 36, No. 1, pp. 44–63. https://www.tandfonline.com/doi/abs/10.1080/02680939.2019.1672211

Herodotou, C., Rienties, B., Hlosta, M., Boroowa, A., Mangafa, C. and Zdrahal, Z. 2020. The scalable implementation of predictive learning analytics at a distance learning university: Insights from a longitudinal case study. *The Internet and Higher Education*, Vol. 45, p. 100725. https://www.sciencedirect.com/science/article/abs/pii/S1096751620300014

Hess, F. 2022. *Schools are getting hit hard by cyberattacks. What can they do about it?* Forbes. https://www.forbes.com/sites/frederickhess/2022/10/21/schools-are-getting-hit-hard-by-cyberattacks-what-can-they-do-about-it/?sh=428b66ae7b71 (Accessed 14 August 2023.)

Hillman, V. 2022. *Edtech Procurement Matters: It Needs a Coherent Solution, Clear Governance and Market Standards*. London, Department of Social Policy, London School of Economics and Political Science. https://www.lse.ac.uk/social-policy/Assets/Documents/PDF/working-paper-series/02-22-Hillman.pdf

Hong, S.-H. 2021. *Why transparency won't save us*. Centre for International Governance Innovation. https://www.cigionline.org/articles/why-transparency-wont-save-us (Accessed 14 August 2023.)

Hooper, L., Livingstone, S. and Pothong, K. 2022. *Problems with Data Governance in UK Schools: The Cases of Google Classroom and ClassDojo*. London, Digital Futures Commission and 5Rights Foundation. https://digitalfuturescommission.org.uk/wp-content/uploads/2022/08/Problems-with-data-governance-in-UK-schools.pdf

Hormazábal, G. L. 2022. Chile: Pioneering the protection of neurorights. *The UNESCO Courier*, 2022, No. 1. https://en.unesco.org/courier/2022-1/chile-pioneering-protection-neurorights

Human Rights Watch. 2022. *"How dare they peep into my private life?" Children's rights violations by governments that endorsed online learning during the COVID-19 pandemic*. https://www.hrw.org/report/2022/05/25/how-dare-they-peep-my-private-life/childrens-rights-violations-governments (Accessed 14 August 2023.)

IBM. 2023. *Cost of a Data Breach Report 2023*. Armonk, New York, United States, IBM Corporation. https://www.ibm.com/downloads/cas/E3G5JMBP

Ienca, M. and Malgieri, G. 2022. Mental data protection and the GDPR. *Journal of Law and the Biosciences*, Vol. 9, No. 1. https://academic.oup.com/jlb/article/9/1/lsac006/6564354

International Bioethics Committee of UNESCO (IBC). 2021. *Report of the International Bioethics Committee of UNESCO (IBC) on the Ethical Issues of Neurotechnology*. UNESDOC Digital Library. https://unesdoc.unesco.org/ark:/48223/pf0000378724 (Accessed 14 August 2023.)

International Commission on the Futures of Education. 2021. *Reimagining Our Futures Together: A New Social Contract for Education*. M. De Sousa (ed.). Paris, UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000379707

International Energy Agency. 2023. *Data centres and data transmission networks*. https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks (Accessed 14 August 2023.)

International Telecommunication Union (ITU). 2021. *Digital Skills Insights 2021*. Geneva, ITU Publications. https://academy. International Telecommunication Union.int/sites/default/files/media2/file/21-00668_Digital-Skill-Insight-210831_CSD%20Edits%206_Accessible-HD.pdf

ITU. 2022. *Internet surge slows, leaving 2.7 billion people offline in 2022*. https://www.itu.int/en/mediacentre/Pages/PR-2022-09-16-Internet-surge-slows.aspx (Accessed 14 August 2023.)

ITU. 2023. *Challenges and Opportunities for Sustainable and Effective Connectivity Policies in Brazilian Schools*. https://www.itu.int/hub/publication/d-phcb-conn_pol-2023-01/

Ipsos. 2022. *Trust in the Internet*. https://www.ipsos.com/sites/default/files/ct/news/documents/2022-11/Trust%20in%20the%20Internet%2C%20Nov%202022.pdf

Jones Day. 2019. *Proposed Algorithmic Accountability Act targets bias in artificial intelligence*. https://www.jonesday.com/en/insights/2019/06/proposed-algorithmic-accountability-act (Accessed 14 August 2023.)

Jones, K. M. L., Asher, A., Goben, A., Perry, M. R., Salo, D., Briney, K. A. and Robertshaw, M. B. 2020. "We're being tracked at all times": Student perspectives of their privacy in relation to learning analytics in higher education. *Journal of the Association for Information Science and Technology*, Vol. 71, No. 9, pp. 1044–1054. https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.24358

Jones, N. 2018. *How to stop data centers from gobbling up the world's electricity*. Nature. https://www.nature.com/articles/d41586-018-06610-y (Accessed 14 August 2023.)

Kerr, K., Norris, T. and Stockdale, R. 2007. Data quality information and decision making: A healthcare case study. *ACIS 2007 Proceedings – 18th Australasian Conference on Information Systems*, pp. 1017–1026. https://aisel.aisnet.org/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1112&context=acis2007

Kivinen, K., Horowitz, M. A., Havula, P., Härkönen, T., Kiili, C., Kivinen, E., Pönkä, H., Pörsti, J., Salo, M., Vuorikari, R. and Vahti, J. 2022. *Digital Information Literacy Guide: A Digital Information Literacy Guide for Citizens in the Digital Age*. Helsinki, Faktabaari. https://faktabaari.fi/dil/digital-information-literacy-guide.pdf

Knight, J. and Hannay, T. 2022. Trust in data, and data in trust. *Education Data Futures: Critical, Regulatory and Practical Reflections*, pp. 294–303. https://educationdatafutures.digitalfuturescommission.org.uk/essays/rethinking-data-futures/trust-data-data-trust (Accessed 14 August 2023.)

Koperniak, S. 2017. *Artificial data give the same results as real data — without compromising privacy*. MIT News. https://news.mit.edu/2017/artificial-data-give-same-results-as-real-data-0303 (Accessed 14 August 2023.)

Kuykendall, K. 2022. *Illuminate data breach spreads to fifth state as Oklahoma City notifies parents*. The Journal. https://thejournal.com/Articles/2022/05/17/Illuminate-Data-Breach-Spreads-to-Fifth-State-as-Oklahoma-City-Notifies-Parents.aspx (Accessed 14 August 2023.)

Lauver, M. 2022. *5 cyber threats facing the education sector*. Security. https://www.securitymagazine.com/articles/97796-5-cyber-threats-facing-the-education-sector (Accessed 14 August 2023.)

Law, N., Woo, D., de la Torre, J. and Wong, G. 2018. *A Global Framework of Reference on Digital Literacy Skills for Indicator* 4.4.2. Montreal, Quebec, Canada, UNESCO-UIS. http://uis.unesco.org/sites/default/files/documents/ip51-global-framework-reference-digital-literacy-skills-2018-en.pdf

Lépine, A. G. 2016. *Essays in the Economics of Education in Brazil*. São Paulo, University of São Paulo Press. https://www.teses.usp.br/teses/disponiveis/12/12138/tde-10022017-121034/pt-br.php

Levin, D. A. 2021. *The State of K-12 Cybersecurity: 2020 Year in Review*. EdTech Strategies, LLC and the K12 Security Information Exchange. https://k12cybersecure.com/wp-content/uploads/2021/03/StateofK12Cybersecurity-2020.pdf

Livingstone, S., and Pothong, K. (eds.) 2022. *Education Data Futures: Critical, Regulatory and Practical Reflections*. https://educationdatafutures.digitalfuturescommission.org.uk/

Locatelli, R. 2018. *Education as a Public and Common Good: Reframing the Governance of Education in a Changing Context*. Paris, UNESCO. (UNESCO Education Research and Foresight Working Papers Series, 22). https://unesdoc.unesco.org/ark:/48223/pf0000261614/PDF/261614eng.pdf.multi

Lohr, S. 2018. *Facial recognition is accurate, if you're a white guy*. The New York Times. https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html (Accessed 14 August 2023.)

Lorente, L. M. L., Arrabal, A. A. and Pulido-Montes, C. 2020. The right to education and ICT during COVID-19: An international perspective. *Sustainability*, Vol. 12, No. 21, p. 91. https://www.mdpi.com/2071-1050/12/21/9091

Lubarsky, B. 2017. Technology explainers: Re-identification of "anonymized" data. *Georgetown Law Technology Review*, pp. 202–213. https://georgetownlawtechreview.org/re-identification-of-anonymized-data/GLTR-04-2017/

Ludic Design for Accessibility. n.d. *All for Play, Play for All*. https://www.ludicdesign.org/ (Accessed 14 August 2023.)

Mahendru, P. 2022. *The state of ransomware in education 2022*. Sophos News. https://news.sophos.com/en-us/2022/07/12/the-state-of-ransomware-in-education-2022/ (Accessed 14 August 2023.)

Mandinach, E. and Gummer, E. (eds.) 2021. *The Ethical Use of Data in Education: Promoting Responsible Policies and Practices*. New York, Teachers College Press.

Marr, B. 2021. *How much data do we create every day? The mind-blowing stats everyone should read*. Bernard Marr & Co. https://bernardmarr.com/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/ (Accessed 14 August 2023.)

Mayer-Schonberger, V. and Cukier, K. 2013. *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London, John Murray.

McKinsey and Company. 2023. *Investing in the rising data center economy*. McKinsey & Company. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/investing-in-the-rising-data-center-economy (Accessed 14 August 2023.)

McKinsey Global Institute. 2021. *Financial data unbound: The value of open data for individuals and institutions*. McKinsey & Company. https://www.mckinsey.com/industries/financial-services/our-insights/financial-data-unbound-the-value-of-open-data-for-individuals-and-institutions (Accessed 14 August 2023.)

McStay, A. 2020. Emotional AI and EdTech: Serving the public good? *Learning, Media and Technology*, Vol. 45, No. 3, pp. 270–283. https://www.tandfonline.com/doi/abs/10.1080/17439884.2020.1686016

Merk, J., Jakobsson, L., Smart, S., Ronsse, S., Popov, D., Björnsson, O., Pawlicki, P. and Hintz, J. 2021. B. Plantenga (ed.), *Human Rights Risks in the ICT Supply Chain*. Make ICT Fair. https://www.ed.ac.uk/files/atoms/files/human_rights_risks_in_the_ict_supply_chain_0.pdf

Metz, R. 2020. *There's a new obstacle to landing a job after college: Getting approved by AI*. CNN Business. https://edition.cnn.com/2020/01/15/tech/ai-job-interview/index.html (Accessed 14 August 2023.)

Miao, F., Hinostroza, J. E., Lee, M., Isaacs, S., Orr, D., Senne, F., Martinez, A.-L., Song, K.-S., Uvarov, A., Holmes, W. and de Dios, B. V. 2022. *Guidelines for ICT in Education Policies and Masterplans*. Paris, UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000380926

Microsoft. 2019. *Ludic Design for Accessibility*. Microsoft Research. https://www.microsoft.com/en-us/research/project/ludicdesign/ (Accessed 14 August 2023.)

MIT Technology Review Insights. 2022. *Global Cloud Ecosystem Index 2022*. https://www.technologyreview.com/2022/04/25/1051115/global-cloud-ecosystem-index-2022/ (Accessed 14 August 2023.)

Mukherjee, S. 2022. *Microsoft launches 'sovereign' cloud for governments*. Reuters. https://www.reuters.com/technology/microsoft-launches-sovereign-cloud-governments-2022-07-19/ (Accessed 14 August 2023.)

Murray, P. 2023. *The SDPC publish the first ever international education sector specific data security standard – The Global Education Security Standard (GESS)*. Access 4 Learning Community. https://home.a4l.org/the-sdpc-publish-the-first-ever-international-education-sector-specific-data-security-standard-the-global-education-security-standard-gess/ (Accessed 14 August 2023.)

Murthy, V. and Ramakrishna, S. 2022. A review on global E-waste management: Urban mining towards a sustainable future and circular economy. *Sustainability*, Vol. 14, No. 2, p. 647. https://www.mdpi.com/2071-1050/14/2/647

Namit, K. and Thi Mai, T. 2019. *Digital school census in 10 weeks? How it was done in Sierra Leone*. World Bank Blogs. https://blogs.worldbank.org/education/digital-school-census-10-weeks-how-it-was-done-sierra-leone (Accessed 14 August 2023.)

National Digital Education Architecture. 2021. *Vidya Samiksha Kendra*. Government of India, Ministry of Education, Department of School Education & Literacy. https://www.ndear.gov.in/vidya-sameeksha-kendra.html (Accessed 14 August 2023.)

National Initiative for Cybersecurity Careers and Studies. 2022. *Workforce Framework for Cybersecurity (NICE Framework)*. Cybersecurity and Infrastructure Security Agency. https://niccs.cisa.gov/workforce-development/nice-framework (Accessed 14 August 2023.)

NITI Aayog. 2018. *National Strategy for Artificial Intelligence: #AIForAll*. https://punjab.gov.in/wp-content/uploads/2023/07/NationalStrategy-for-AI.pdf (Accessed 21 August 2023.)

ODI Learning. 2020. *The Data Skills Framework*. https://learning.theodi.org/courses/data-skills-framework (Accessed 14 August 2023.)

Odhiambo, R. A., Wakoli, E. and Rodrot, M. 2022. *Africa's Ed-Tech Platforms: Protecting Children's Right to Privacy*. Center of Intellectual Property and Technology Law. https://www.africanobservatory.ai/ai4d-resources/data-privacy-in-africas-ed-tech-platforms-childrens-right-to-privacy

Organisation for Economic Co-operation and Development (OECD). 2016. *Understanding Social Impact Bonds*. https://www.oecd.org/cfe/leed/UnderstandingSIBsLux-WorkingPaper.pdf

OECD. 2019. *OECD Recommendation on Responsible Innovation in Neurotechnology*. https://www.oecd.org/science/recommendation-on-responsible-innovation-in-neurotechnology.htm (Accessed 14 August 2023.)

OECD. 2020. *Mapping Approaches to Data and Data Flows: Report of the G20 Digital Economy Task Force – Saudi Arabia 2020*. https://www.oecd.org/sti/mapping-approaches-to-data-and-data-flows.pdf

OECD. 2021*a*. *Enhancing Data Informed Strategic Governance in Education in Estonia*. Paris, OECD Publishing. https://www.oecd-ilibrary.org/education/enhancing-data-informed-strategic-governance-in-education-in-estonia_11495e02-en

OECD. 2021*b*. *Public Governance in Costa Rica*. https://www.oecd.org/governance/costa-rica-public-governance-evaluation-accession-review.pdf

Ofqual. 2020. *Statement from Roger Taylor, Chair, Ofqual: How grades for GCSE, AS, A level, Extended Project Qualification and Advanced Extension Award in maths will be awarded this summer*. GOV.UK. https://www.gov.uk/government/news/statement-from-roger-taylor-chair-ofqual (Accessed 14 August 2023.)

OneTrust Data Guidance. 2022. *Costa Rica - Data Protection Overview*. https://www.dataguidance.com/notes/costa-rica-data-protection-overview (Accessed 14 August 2023.)

Pangrazio, L. 2016. Reconceptualising critical digital literacy. *Discourse: Studies in the Cultural Politics of Education*, Vol. 37, No. 2, pp. 163–174. https://www.tandfonline.com/doi/abs/10.1080/01596306.2014.942836

Perrigo, B. 2022. *Inside Facebook's African sweatshop*. Time. https://time.com/6147458/facebook-africa-content-moderation-employee-treatment/ (Accessed 14 August 2023.)

Quinton, S. and Reynolds, N. 2018. Digital Research as a Phenomenon and a Method. M. Waters (ed.), *Understanding Research in the Digital Age*. London, SAGE Publications Ltd.

Ravitch, D. 2010. *The myth of charter schools*. Brookings. https://www.brookings.edu/articles/the-myth-of-charter-schools/ (Accessed 14 August 2023.)

Razquin, P., Strath, A. and Kosbar, Y. 2023. *A Review of Alternative Country Models and Strategies for Financing Digital Learning*. UNESDOC Digital Library. https://unesdoc.unesco.org/ark:/48223/pf0000386086 (Accessed 14 August 2023.)

Regan, P. M. and Khwaja, E. T. 2019. Mapping the political economy of education technology: A networks perspective. *Policy Futures in Education*, Vol. 17, No. 8, pp. 1000–1023. https://journals.sagepub.com/doi/full/10.1177/1478210318819495

Reinsel, D., Gantz, J. and Rydning, J. 2018. *The Digitization of the World: From Edge to Core*. Framingham, Massachusetts, United States, International Data Corporation (IDC). https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

Re-use and Recycling EU Social Enterprises network (RREUSE). 2017. *Reduced taxation to support re-use and repair*. https://www.rreuse.org/wp-content/uploads/RREUSE-position-on-VAT-2017-Final-website_1.pdf

Rommelfanger, K. S., Pustilnik, A. and Salles, A. 2022. *Mind the gap: Lessons learned from neurorights*. AAAS Science & Diplomacy. https://www.sciencediplomacy.org/article/2022/mind-gap-lessons-learned-neurorights (Accessed 14 August 2023.)

Sanders, C. 2022. *Microsoft Cloud for Sovereignty: The most flexible and comprehensive solution for digital sovereignty*. Official Microsoft Blog. https://blogs.microsoft.com/blog/2022/07/19/microsoft-cloud-for-sovereignty-the-most-flexible-and-comprehensive-solution-for-digital-sovereignty/ (Accessed 14 August 2023.)

Sattler, S. and Pietralla, D. 2022. Public attitudes towards neurotechnology: Findings from two experiments concerning Brain Stimulation Devices (BSDs) and Brain-Computer Interfaces. *PLoS One*, Vol. 17, No. 11, p. e0275454. https://pubmed.ncbi.nlm.nih.gov/36350815/

Schneider, A. C., Arcego, E. and Schneider, A. C. 2019. The University for All Program (ProUni): From statistics up to the law and development approach. *Planejamento e Políticas Públicas*, No. 52, pp. 409–436. https://repositorio.ipea.gov.br/bitstream/11058/9760/1/ppp_n52_university.pdf

Selwyn, N. 2015. Data entry: Towards the critical study of digital data and education. *Learning, Media and Technology*, Vol. 40, No. 1, pp. 64–82. https://www.tandfonline.com/doi/full/10.1080/17439884.2014.921628

Sheridan, M. P. and Rowsell, J. 2010. *Design Literacies: Learning and Innovation in the Digital Age, Literacies* (gen. ed. D. Barton). London, Routledge.

Singer, N. 2022. *A cyberattack illuminates the shaky state of student privacy*. The New York Times. https://www.nytimes.com/2022/07/31/business/student-privacy-illuminate-hack.html (Accessed 14 August 2023.)

Singer, N. 2023. *In classrooms, teachers put A.I. tutoring bots to the test*. The New York Times. https://www.nytimes.com/2023/06/26/technology/newark-schools-khan-tutoring-ai.html (Accessed 14 August 2023.)

Softtek. 2021. Data minimalism: A new philosophy in the era of big data. The Softtek Blog. https://blog.softtek.com/en/data-minimalism-a-new-philosophy-in-the-era-of-big-data (Accessed 14 August 2023.)

South African Department of Basic Education. 2022. *Thutong Portal: Delivering information, curriculum, and support materials to the South African schooling community.* Thutong. https://www.thutong.doe.gov.za/

Stahl, W. M. and Karger, J. 2016. Student data privacy, digital learning, and special education: Challenges at the intersection of policy and practice. *Journal of Special Education Leadership*, Vol. 29, No. 2, pp. 79–88. https://files.eric.ed.gov/fulltext/EJ1118549.pdf

Standaert, M. 2019. *Chinese primary school halts trial of device that monitors pupils' brainwaves*. The Guardian. https://www.theguardian.com/world/2019/nov/01/chinese-primary-school-halts-trial-of-device-that-monitors-pupils-brainwaves (Accessed 14 August 2023.)

Statista. 2022. *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025*. https://www.statista.com/statistics/871513/worldwide-data-created/ (Accessed 14 August 2023.)

Statista. 2023. *The 100 largest companies in the world by market capitalization in 2023*. https://www.statista.com/statistics/263264/top-companies-in-the-world-by-market-capitalization/ (Accessed 14 August 2023.)

Synced. 2020. *Machine unlearning: Fighting for the right to be forgotten*. SyncedReview. https://medium.com/syncedreview/machine-unlearning-fighting-for-the-right-to-be-forgotten-c381f8a4acf5 (Accessed 14 August 2023.)

Tarrant, D. 2021. *Data literacy: What is it and how do we address it at the ODI?* Open Data Institute. https://theodi.org/article/data-literacy-what-is-it-and-how-do-we-address-it-at-odi/ (Accessed 14 August 2023.)

Tidey, A. 2019. *EU e-waste 'illegally' exported to developing countries: Report*. Euronews. https://www.euronews.com/my-europe/2019/02/07/eu-e-waste-illegally-exported-to-developing-countries-report (Accessed 14 August 2023.)

Turner, S. 2022. *Reality check on technology uses in UK state schools*. Digital Futures Commission, 5Rights Foundation. https://digitalfuturescommission.org.uk/blog/reality-check-on-technology-uses-in-uk-state-schools/ (Accessed 14 August 2023.)

Turner, S., Pothong, K. and Livingstone, S. 2022. *Education data reality: The challenges for schools in managing children's education data – Digital Futures Commission June 2022*. Digital Futures Commission, 5Rights Foundation. https://digitalfuturescommission.org.uk/wp-content/uploads/2022/06/Education-data-reality-report.pdf

United Nations Educational, Scientific and Cultural Organization (UNESCO). 2017. *Global Education Monitoring Report 2017/8: Accountability in Education: Meeting Our Commitments*. Paris, UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000259338

UNESCO. 2019. *Beijing Consensus on Artificial Intelligence and Education*. Paris, UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000368303

UNESCO. 2020. *UNESCO Qualifications Passport*. https://www.unesco.org/en/emergencies/qualifications-passport (Accessed 14 August 2023.)

UNESCO. 2022*a*. *K-12 AI Curricula: A Mapping of Government-endorsed AI Curricula*. Paris, UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000380602

UNESCO. 2022*b*. *Minding the Data: Protecting Learners' Privacy and Security*. Paris, UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381494

UNESCO. 2022*c*. *Re-Imagining the Future of Education Management Information Systems: Ways Forward to Transform Education Data Systems to Support Inclusive, Quality Learning for All*. Paris, UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381618

UNESCO. 2022*d*. *The RewirEd Global Declaration on Connectivity for Education: #ConnectivityDeclaration*. Paris, UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381482

UNESCO. 2023*a*. *Enhancing TVET Through Digital Transformation in Developing Countries*. Paris, UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000385988

UNESCO. 2023*b*. *Global Education Monitoring Report 2023: Technology in Education: A Tool on Whose Terms?* Paris, UNESCO. https://unesdoc.unesco.org/ark:48223/pf0000385723

UNESCO. 2023*c*. *Recommendation on the Ethics of Artificial Intelligence*. https://en.unesco.org/artificial-intelligence/ethics (Accessed 14 August 2023.)

UNESCO. 2023*d*. *Transforming Education Together: The Global Education Coalition in Action*. Paris, UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000384812

UNESCO. n.d. *Open Data*. https://www.unesco.org/en/open-solutions/open-data (Accessed 14 August 2023.)

UNESCO and Dubai Cares. 2021. *Rewired Global Declaration on Connectivity for Education*. https://unesdoc.unesco.org/ark:/48223/pf0000380598/PDF/380598eng.pdf.multi

UNESCO Institute for Lifelong Learning. 2022. *5th Global Report on Adult Learning and Education: Citizenship education: Empowering Adults for Change*. Hamburg, Germany, UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381666

UNESCO Institute for Statistics (UIS). 2020. *Data Innovation for Producing SDG 4 Indicators: An EMIS Metadata Global Analytical Report*. Montreal, Quebec, Canada, UNESCO-UIS. http://uis.unesco.org/sites/default/files/documents/ip65-emis_typology-final_en2.pdf

UIS. 2020. *Data Innovation for Producing SDG 4 Indicators: A Global Analytical Report*. Montreal, Quebec, Canada, UNESCO-UIS. https://unesdoc.unesco.org/ark:/48223/pf0000374784

UIS. 2021. *Learning Data Compact – UNESCO, UNICEF, and the World Bank Unite to End the Learning Data Crisis*. https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2021/06/Learning-Data-Announcement.pdf

UNESCO and the Collective Consultation of NGOs. 2015. *2015 NGO Forum Declaration: Towards the Right to Inclusive Quality Public Education and Lifelong Learning Beyond 2015*. UNESDOC Digital Library. https://unesdoc.unesco.org/ark:/48223/pf0000233243 (Accessed 14 August 2023.)

United Nations Conference on Trade and Development (UNCTAD). 2021. *Digital Economy Report 2021: Cross-border Data Flows and Development: For Whom the Data Flow*. New York, United Nations Publications. https://unctad.org/page/digital-economy-report-2021

UNCTAD. 2023. *G20 Members' Regulations of Cross-Border Data Flows*. Geneva, United Nations. https://unctad.org/system/files/official-document/dtlecdc2023d1_en.pdf

UNCTAD. n.d. *Data Protection and Privacy Legislation Worldwide*. https://unctad.org/page/data-protection-and-privacy-legislation-worldwide (Accessed 14 August 2023.)

Unitas Financial Services. n.d. *The Use of AI in Loan Decisions - Unitas Financial Services*. https://www.unitas360.com/blog/the-use-of-ai-in-loan-decisions (Accessed 14 August 2023.)

United Nations. 1989. *Convention on the Rights of the Child*. https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child

United Nations. 2020. *Road Map for Digital Cooperation: Implementation of the Recommendations of the High-level Panel on Digital Cooperation*. https://www.un.org/en/content/digital-cooperation-roadmap/

United Nations. 2021. *Our Common Agenda: Report of the Secretary-General*. New York, United Nations. https://www.un.org/en/content/common-agenda-report/

United Nations Office of the Secretary-General's Envoy on Technology. 2023. *Global Digital Compact*. United Nations. https://www.un.org/techenvoy/global-digital-compact (Accessed 14 August 2023.)

United States Securities and Exchange Commission. 2021. *SEC charges Pearson Plc for misleading investors about cyber breach*. SEC. gov. https://www.sec.gov/enforce/33-10963-s (Accessed 14 August 2023.)

Van Wetering, M., Booij, E. and van Bruggen, W. 2020. S. van Dort and Schlingemann (eds.), *Education in an Artificially Intelligent World: Kennisnet Technology Compass 2019-2020*. Zoetermeer, Kingdom of the Netherlands. https://www.slideshare.net/eraser/education-in-an-artificially-intelligent-world-kennisnet-technology-compass-20192020

Veale, M., Binns, R. and Edwards, L. 2018. Algorithms that remember: Model inversion attacks and data protection law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 376. https://royalsocietypublishing.org/doi/10.1098/rsta.2018.0083

Vereecken, W., Van Heddeghem, W., Colle, D., Pickavet, M. and Demeester, P. 2010. *Overall ICT Footprint and Green Communication Technologies*. https://www.researchgate.net/publication/224137694_Overall_ICT_footprint_and_green_communication_technologies

Veritas Technologies. n.d. *The Definitive Guide to Cloud Storage Pricing*. Veritas. https://www.veritas.com/information-center/the-definitive-guide-to-cloud-storage-pricing (Accessed 14 August 2023.)

Vision Empower Trust. n.d. Vision Empower. http://visionempowertrust.in/ (Accessed 14 August 2023.)

Vuorikari, R., Kluzer, S. and Punie, Y. 2022. *DigComp 2.2: The Digital Competence Framework for Citizens – With New Examples of Knowledge, Skills and Attitudes*. Luxembourg, Publications Office of the European Union. https://publications.jrc.ec.europa.eu/repository/handle/JRC128415

Walonoski, J., Kramer, M., Nichols, J., Quina, A., Moesel, C., Hall, D., Duffett, C., Kudakwashe, D., Gallagher, T. and McLachlan, S. 2017. Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *Journal of the American Medical Informatics Association*, Vol. 25, No. 3, pp. 230–238. https://academic.oup.com/jamia/article/25/3/230/4098271

Whittlestone, J., Nyrup, R., Alexandrova, A. and Cave, S. 2019. The role and limits of principles in AI ethics: Towards a focus on tensions. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 195–200. https://dl.acm.org/doi/10.1145/3306618.3314289

Williamson, B. 2016. Digital education governance: Data visualization, predictive analytics, and 'real-time' policy instruments. *Journal of Education Policy*, Vol. 31, No. 2, pp. 123–141. https://www.tandfonline.com/doi/full/10.1080/02680939.2015.1035758

World Bank. 2018. *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC, World Bank. https://www.worldbank.org/en/publication/wdr2018

World Bank. 2023 (Forthcoming). *Innovative Financing of Education Technology: As Part of Maximizing Financing for Development*.

Yao, H., Brossard, M., Mizunoya, S., Nasir, B., Walugembe, P., Cooper, R., Rafique, A. and Reuge, N. 2021. *How Much Does Universal Digital Learning Cost?* Florence, Italy, UNICEF Office of Research – Innocenti. (Policy Brief.) https://www.unicef-irc.org/publications/1301-how-much-does-universal-digital-learning-cost.html

Zhao, J. 2022. Call for a new data governance structure for datafied childhood. *Education Data Futures: Critical, Regulatory and Practical Reflections*, pp. 314–327. https://cms.educationdatafutures.digitalfuturescommission.org.uk/wp-content/uploads/2022/09/Education-Data-Futures_Essay-22.pdf

Zhu, M. 2020. *An algorithmic jury: Using artificial intelligence to predict recidivism rates*. Yale Scientific. https://www.yalescientific.org/2020/05/an-algorithmic-jury-using-artificial-intelligence-to-predict-recidivism-rates (Accessed 14 August 2023.)

Zuboff, S. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* London, Profile Books Limited.

Annex

United Nations Conference on Trade and Development (UNCTAD) survey on policy approaches to cross-border data flows with trust

Unilateral approaches	
Adequacy, standard contractual clauses or binding corporate rules	Provides an assessment of data protection (for personal data) before allowing data flows to take place.
Data transfer plans requiring government approval	Employed by countries such as Indonesia with respect to geospatial data, health data and personal data protection.
Laws and regulations requiring specific consent	Consent from the data subject as a condition for international data flows. Often used in the context of personal data.
Data localization	Limits or conditions data to flow freely. For example, personal data generated in the Russian Federation and by its inhabitants must be stored locally first before it can be transferred abroad. In the context of cloud computing, Brazil requires a recent data backup to be maintained in its territory. Saudi Arabia requires data and data infrastructure for cloud and Internet of Things (IoT) systems to be maintained within the country.[21] The same holds for the Republic of Korea's cloud computing services for public procurement, for which data need to be stored in the country. Türkiye has a provision that requires data from the financial sector and capital markets to have their primary and secondary versions stored domestically; beyond this there are no explicit restrictions on data flows.
Bilateral approaches	
Contracts and Memoranda of Understanding for international parties to enable data flows	Predominantly for personal data as well as data linked to commodities trading (e.g. Canada, Mexico, Singapore and the United States of America). Similar to the above, they put the responsibility on the local data controller to ensure that non-domestic data processors adhere to relevant safeguards.
Judicial cooperation	For example, in Argentina, Canada, the Russian Federation, the United Kingdom of Great Britain and Northern Ireland and the United States.
Regulatory cooperation	Enhanced international regulatory cooperation is secured through arrangements between regulators on issues such as enforcement, and through dialogues and agreements between governments. Regulatory cooperation is also a key area mentioned in the 2021 G7 Roadmap for Cooperation on Data Free Flow with Trust.[22]

21 See the cloud computing regulatory framework, the IoT regulatory framework, and the essential cybersecurity controls.

22 See https://www.gov.uk/government/publications/g7-digital-and-technology-ministerial-declaration.

Multilateral frameworks

General Data Protection Regulation (GDPR): The GDPR, applicable in the European Union (EU), is a comprehensive data protection regulation applicable in the EU and European Economic Area (EEA), which sets rules and standards for the protection of personal data and governs cross-border transfers of personal data to countries outside the EU/EEA.

Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108): The Council of Europe's Convention 108 is an international treaty that aims to protect the privacy and fundamental rights of individuals regarding the processing of personal data. It covers both private and public sectors and applies to any personal data held or processed by automated means, and it establishes principles for the fair and lawful processing of personal data. It emphasizes the need for informed consent, purpose limitation, data quality, data security, transparency and accountability. Convention 108 has been supplemented by additional protocols to address emerging challenges in data protection, such as the Protocol on Transborder Data Flows and the Protocol on Supervisory Authorities and Transborder Data Flows. It is important to note that Convention 108 has also undergone a modernization process to adapt to new technological developments and legal frameworks. The modernized Convention, known as Convention 108 +, was adopted in 2018 to enhance data protection and address emerging challenges in the digital era.

European Union directives and regulations on establishing a digital single market, with an aim of fostering the free flow of data within the EU Member States.

Multiple initiatives by the United Nations, including the Secretary-General's "Our Common Agenda" and the United Nation's Children's Fund's (UNICEF) Global Development Commons, as well as initiatives from civil society such as the Datasphere Initiative.

African Union Data Policy Framework

G7 Trade Ministers' Digital Trade Principles

Asia-Pacific Data Privacy Framework, developed by the Asia-Pacific Economic Cooperation (APEC) which outlines principles and implementation guidance for member economies to facilitate secure and seamless cross-border data flows while safeguarding individual privacy. The framework also promotes accountability, transparency and cooperation among participating economies.

Joint Initiative Negotiations on E-commerce, the Comprehensive and Progressive Agreement for Trans-Pacific Partnership, the Regional Comprehensive Economic Partnership and the United States-Mexico-Canada Agreement.

The United Nations Educational, Scientific and Cultural Organization (UNESCO) Recommendation on Open Science, adopted in 2021, and the UNESCO Recommendation on Open Educational Resources, adopted in 2019, represent frameworks and guidance for collaboration and information exchange between countries.

The Organisation for Economic Co-operation and Development (OECD) Guidelines on the Protection of Privacy and Transborder Flows of Personal Data: These guidelines provide recommendations for member coun-tries to ensure the privacy and security of personal data during cross-border transfers.

General trends	
Lead agencies for data governance and whole-of-government approaches	Whole-of-government approaches that clearly identify one lead organization for the government's strategy on data governance may help coordinate requirements and demands from society and economy. For example, in the United Kingdom, this responsibility is shared between two entities. The Department for Digital, Culture, Media and Sport leads on general data governance frameworks, while the Central Digital and Data Office leads on government data.
Multi-stakeholder coordination	Canada introduced its Digital Charter following multi-stakeholder consultations. A Digital Charter Implementation Act is currently being discussed in Parliament. Mexico introduced the *Abramos México* initiative in February 2022, which aims to develop a National Open Data Policy in a public, open and collaborative way with multi-sector inputs. This endeavour is promoted by multiple agencies (including the National Institute for Transparency, Access to Information and Protection of Personal Data, the National Transparency System, and the regulatory body for statistical and geographical information), civil society and academia. The Republic of Korea has set up an Open Data Strategy Council under the Prime Minister, with participation of other ministers and co-chaired by the private sector. The Council guides and coordinates the Government's policies, plans on open government data and monitors their implementation. It is led by the Ministry of Interior and Safety. The United Kingdom incorporates multi-stakeholder feedback through various mechanisms and fora, including the National Data Strategy Forum and the International Data Transfers Expert Council, expert and advisory groups, and expert networks from academia and the private sector. Furthermore, the Government publishes public and stakeholder consultations for policy interventions linked to data.

Source: Partly compiled from mechanisms for multi-stakeholder dialogue on data governance and a survey carried out by UNCTAD in the G20 Member States and invited guests under the Indonesian Presidency in 2022.

BROADBAND COMMISSION
FOR SUSTAINABLE DEVELOPMENT