aapti institute

moz://a

USAID
FROM THE AMERICAN PEOPLE

Strengthening Data Ecosystems in Indian Schools

This is a work of independent research, commissioned by **Mozilla Foundation** in 2023-2024 and supported by a grant from **USAID**.

It was produced by **Aapti Institute**, a public research institute that works on the intersection of technology and society. Aapti examines the ways in which people interact and negotiate with technology both offline and online.

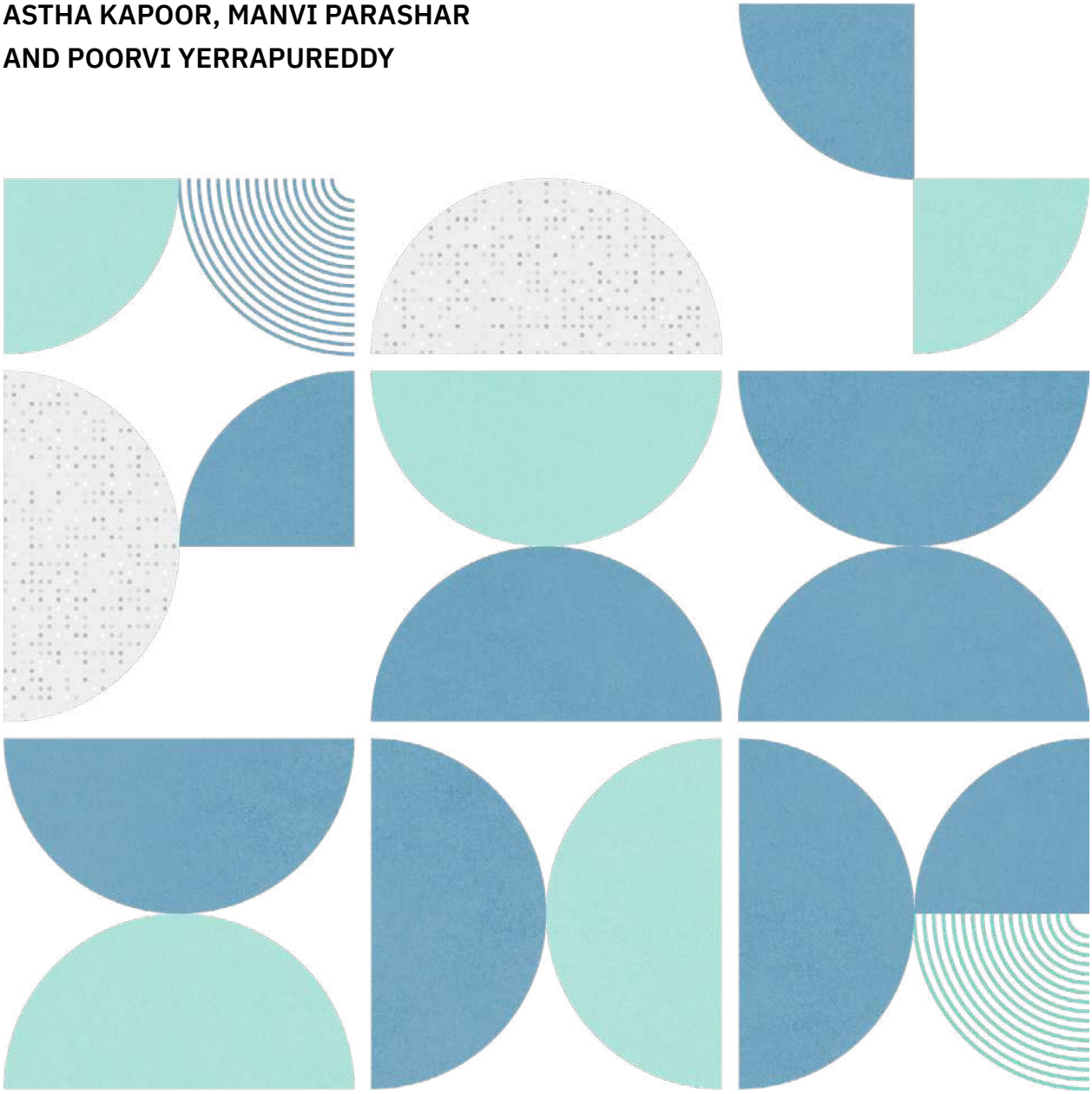
ACKNOWLEDGEMENTS

In addition to contribution from the wider Aapti team, this report also draws from the expertise of numerous academic and industry experts and practitioners, including Dr Ganesh Nigam, Education Specialist at UNICEF; Roshan Nair, Digital Inclusion Specialist at USAID; Karthik Menon, leading Government Partnerships in Education at Pratham Education Foundation; Archana Sivasubramanian, leading the Centre for Technology Law and Policy at Sai University; and Dr Amit Chandra from the Centre for Civil Society. We are grateful for their input during our workshop and conversations. A heartfelt thank you goes out to Quicksand Design Studios for lending their expertise for the workshop, their contribution to the exercises and the sessions are invaluable. Lastly, we extend our deepest gratitude to all the members of the working groups. Your active participation and dedication have been the cornerstone of this initiative.

Report design: Meher Rajpal & Antara Madavane | **Cover illustration:** Teesta Nayak

Strengthening Data Ecosystems in Indian Schools

ASTHA KAPOOR, MANVI PARASHAR
AND POORVI YERRAPUREDDY



aapti institute
moz://a
USAID
FROM THE AMERICAN PEOPLE

Table of Contents

SECTION 1

Introduction

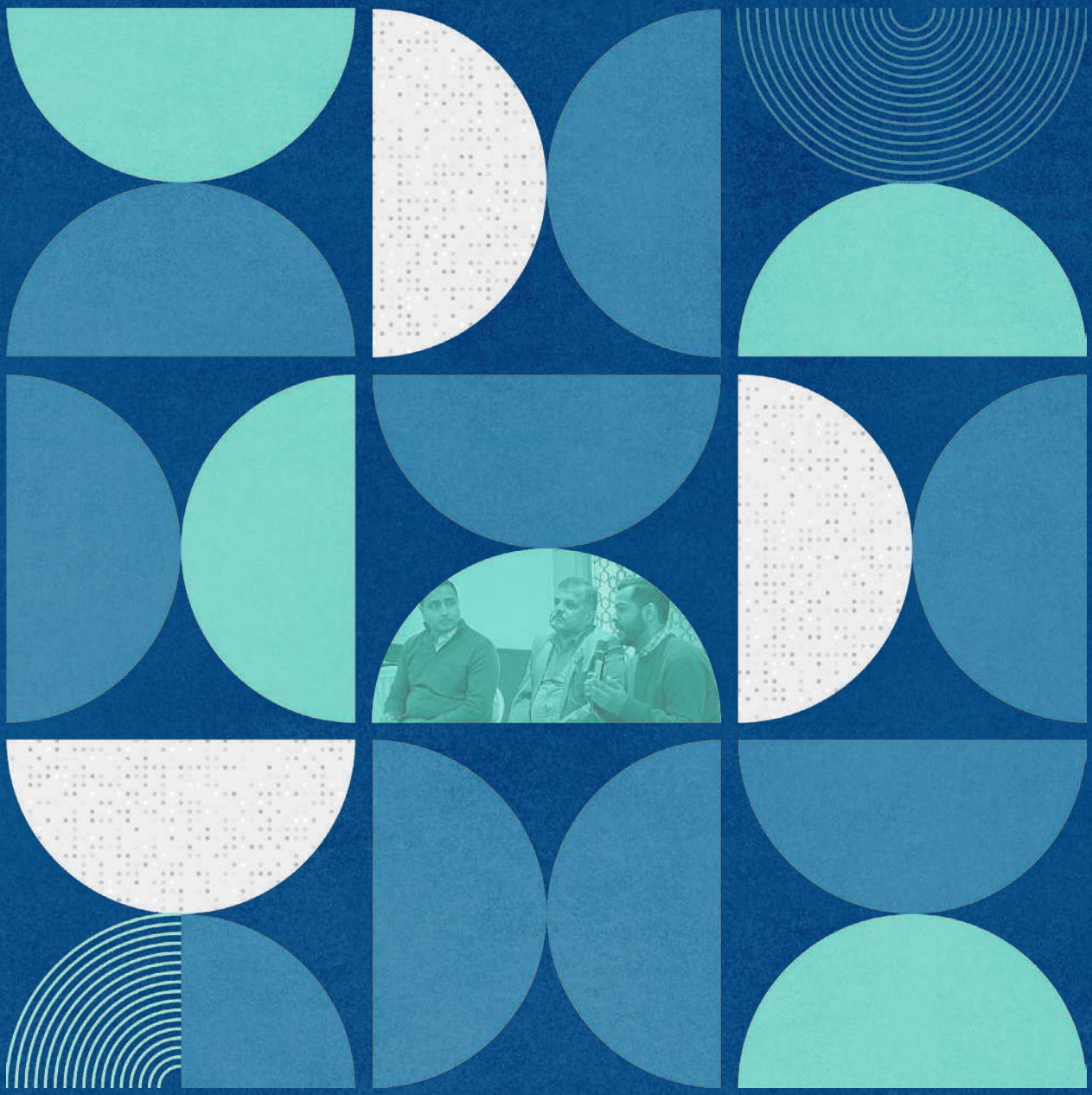
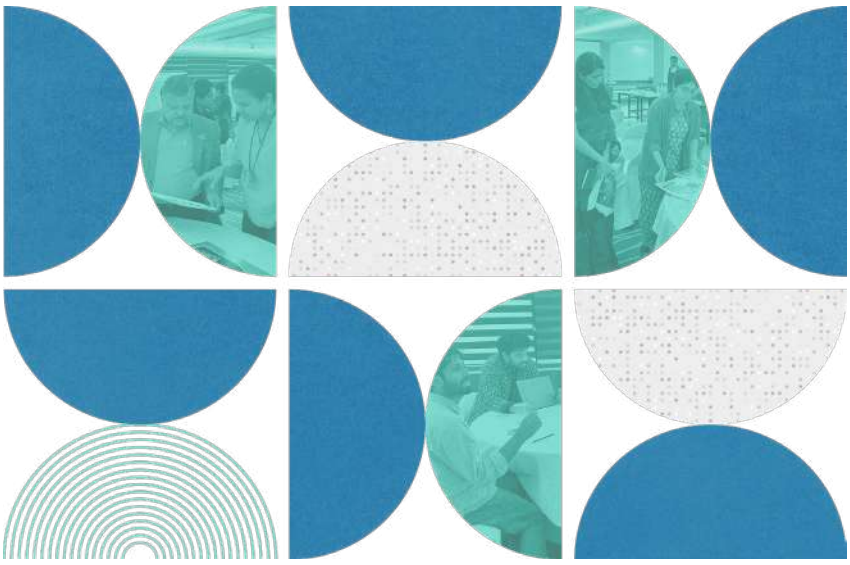


TABLE OF CONTENTS ANNEXURES

SECTION 1

Introduction

The Indian school education system, one of the world's largest, includes over 1.49 million schools, with 335,726 of those being private schools[1], about 9.5 million teachers, and serves more than 265 million students from various socio-economic backgrounds across 28 States and 8 Union Territories[2]. In this landscape, data plays a pivotal role in shaping policies and interventions. School education data, such as enrollment rates, student-teacher ratios, attendance, assessment results, socio-economic data and data related to school infrastructure is utilised for framing such policies and interventions. India is increasingly using technology in various sectors such as healthcare, manufacturing, agriculture and financial services delivery and recognises the need for a similar digital infrastructure in education.



Integrating data and technology in education can lead to personalised learning, improved educational outcomes, equitable access to high-quality education for all students, and efficient

[1] Schools Management Report UDISE+ Dashboard
[2] Unified District Information System for Education +

allocation of resources. However, challenges such as the scale and varied nature of the education system, infrastructure gaps, and the need for regular training on new technologies pose challenges in the standardisation of data collection, efficient use of the data collected, and the equitable distribution of resources. These issues should be considered and addressed to ensure that any policy formulated bridges the gaps in education, rather than widening them.

The three elements central to data and technology in the Indian education ecosystem at present are- National Education Policy 2020[3] (NEP), The National Digital Education Architecture (NDEAR)[4] created under the NEP, which has set foundational principles, standards and specifications for building a robust digital education infrastructure and finally, the Unified District Information System for Education Plus (UDISE+)[5], which manages school-level data across

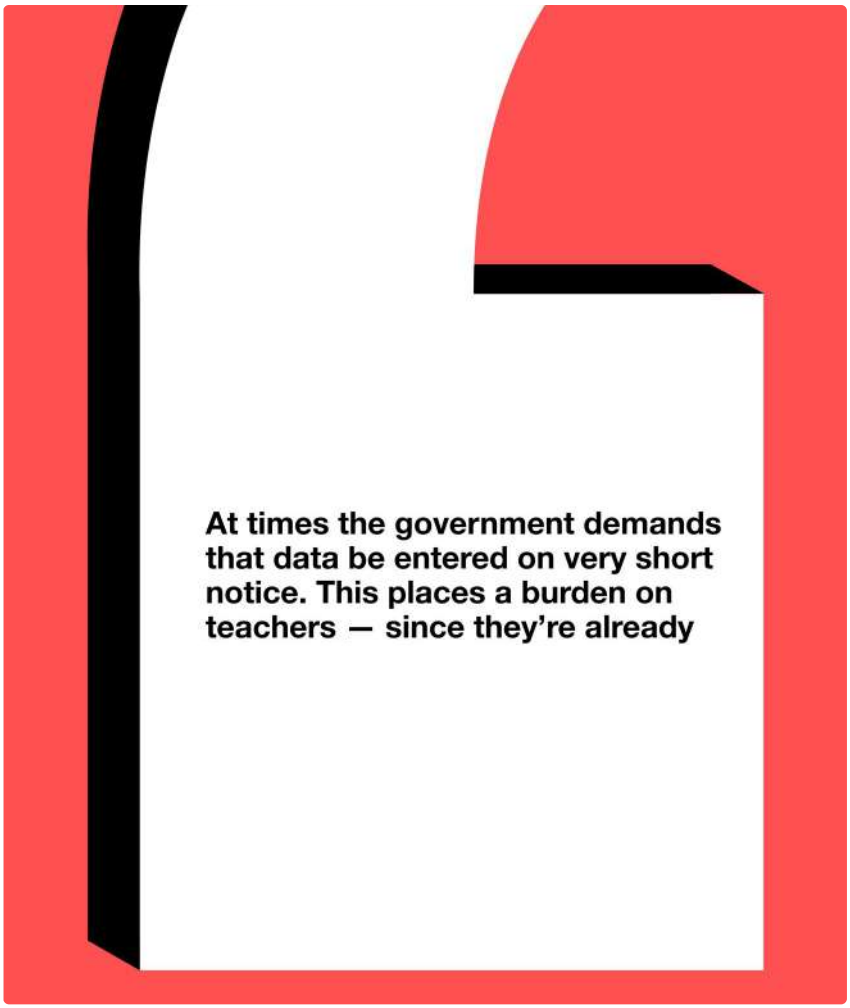


Image courtesy of Quicksand Design Studio.

[3] Ministry of Human Resource Development, Government of India 'National Education Policy, 2020'

[4] NDEAR-Open Standards_Version.pdf Pages 15-18 *"NDEAR envisions digital infrastructure for school education that allows all ecosystem actors across government, civil society, and market players to build and innovate platforms, solutions and tools that are compatible or compliant with NDEAR and the 12 NDEAR Architecture Principles"*

[5] Unified District Information System for Education +

the country. However, in relation to the data made available by UDISE+ the pre-primary data is not representative of all pre-primary schools in the country. Specifically, the system does not include data from stand alone pre-primary schools and the pre-primary enrolment of the Early Childhood Care and Education (ECCE)[6] system, run by the Ministry of Women and Child Welfare. This segment of data on children between the ages 3 to 6 is overseen by the Anganwadi Centres (AWC) through the POSHAN[7] Tracker[8]. The POSHAN tracker classifies children under this age group into two types: as school going[9] (those who attend school other than Anganwadi Centres and non school going (those who attend the ECCE system under Anganwadi Centre). At present, this data on children between the ages 3 to 6 managed through the POSHAN tracker is not integrated with UDISE+[10]. This highlights significant silos in the state of data systems in education. The fragmentation of data extends beyond these two platforms, involving the data systems managed by various state governments in the country in contrast to the central UDISE+ system. This lack of integration of data sets creates impediments in efforts to draw meaningful insights and actionable outcomes due to lack of availability, access and visibility of comprehensive data.

The NDEAR provides the foundational framework for a robust digital infrastructure for the education sector in India. With an aim to create a digital framework for school education, enabling various stakeholders like government, civil society, and businesses to

Working group members at the SDE event in New Delhi.

[6] Early Childhood Care and Education (ECCE) and the Foundational Learning framework, released by the Ministry of Women and Child Development, is focused on the needs of the child and works towards improving child care and a developmentally appropriate environment for children. This framework focuses on providing guidelines for child care and early educational practices, including the provision of quality ECCE in pre- schools, training of ECCE teachers, and child-friendly facilities.

[7] Pradhan Mantri Poshan Shakti Nirman (PM POSHAN) scheme is focused on the nutritional well-being of students, a cornerstone of effective learning, and development. Under the Scheme, there is provision of hot cooked meal to children of pre-schools or Bal Vatika (before class I) in primary schools in addition to the 11.80 crore children of classes I to VIII studying in 11.20 lakh schools

[8] POSHAN Tracker Dashboard

[9] Poshan Tracker FAQ

[10] Ministry of Education, Report on UDISE + (2020-21) Report Page 11 *"It may be noted that the pre-primary data contained in the UDISE+ system does not include the pre-primary enrolment of the Early Childhood Care and Education (ECCE) system run by the Ministry of Women and Child Welfare through the Anganwadi Centers and stand alone pre-primary education centers run by different kindergartens."*

develop and integrate compatible platforms and tools, NDEAR provides 12 guiding principles. Currently, there are few initiatives and projects created under the umbrella of NDEAR, which have been implemented, one such initiative such as the Vidya Samiksha Kendra[11] which is a technology enabled initiative to monitor the learning progress of students, across every grade in every school, which has also been recognised as a best practice by the World Bank[12]. Under the NEP, 2020 a National Education Technology Forum (NETF)[13] was established, which is a platform for interaction and sharing of best practices in educational technology and testing of ed-tech products in a sandbox environment to help formulate a learner centric approach. Its mandate is to also develop accessible educational software with an aim to improve teaching, learning, and administrative efficiency in education. Through the NETF, a framework for Education Ecosystem Registries (EER)[14] has been created. These electronic registries have been envisioned to create a high trust low cost infrastructure enabling ease of access to data, and to allow users and entities to manage their credentials. These could include details of students, faculty, educational institutes, skill providers, and edtech companies. This is expected to be a force multiplier for the existing systems which operate in silos leading to lack of visibility and access to data, which result in the absence of data-driven insights necessary for effective policy making.

These initiatives undertaken to rapidly adopt technology bring us to the forefront of a crucial challenge: *ensuring data quality[15], security[16] and data culture[17]*. The possibility of instances of erroneous data entry (for example phone numbers being recorded in place of names) also highlight the systemic issue of data hygiene.

Accurate data profiling can be useful to tackle the data utility challenge. Predictive data analysis not only identifies problems, but also gives insights that enable actionable solutions. Especially for understanding diverse factors like age, socio-economic backgrounds, and gender-specific issues. For instance, many girls tend to leave school after the 10th grade[18] due to various reasons, including insufficient sanitation facilities[19] and distance to higher education institutes. Additionally, societal norms often place household duties and early marriage over education for girls. Effectively segmenting data to identify and support students at risk

[11] Vidya Samiksha Kendra- NDEAR Project; Ministry of Education, Department of School Education and Literacy

[12] World Bank, Learning Recovering to Acceleration Page 55

[13] Ministry of Human Resource Development, Government of India, National Education Policy, 2020 Page 56

[14] Design and Architecture Framework for Education Ecosystem Registry

[15] Pandit, Niraj. "State pressuring school principals to manipulate education data, allege principals and teachers." Hindustan Times, September 2023.

[16] Elliot, Victoria. Mehrotra, Dhruv. "A Major App Flaw Exposed the Data of Millions of Indian Students." Wired, January 2023. ; Rampal, Anita. "Mining Children: Why Dataveillance in the Name of Education or Health Must Stop." The Wire, May 2023.

[17] Jhingran, Dhir. "Data collection alone cannot improve learning outcomes in State-run schools." Hindustan Times, November 2016.

[18] "Dropout rate of girls after Class X still worrisome." The Times of India, January 2022.

[19] Screwala, Zarina. "How building school toilets can keep our girls in school." Forbes India, February 2023. ; "School dropout rate of girls sharply declined due to schemes for separate toilets, sanitary pads: Prez." The Print, January 2023.

of dropping out is crucial, and this can be done by accurate predictive analysis of data[20].

Predictive tools can help foresee educational challenges, such as high dropout rates, high academic failures in specific subjects like mathematics. By analysing historical data, we can understand these issues. However, the success of such an analysis will depend on the quality of the data. The Data, Information, Knowledge, Insight, Wisdom (DIKIW) Pyramid[21] highlights the importance of every step in data processing, emphasising that high-quality data is crucial for deriving meaningful insights.

There are broader, systemic challenges in the Indian education system. The long-term nature of educational outcomes often clashes with the immediate living conditions of many families. Issues like the lack of integration between different educational and nutritional tracking systems, regional imbalances, and gender disparities require attention. These complexities extend into higher education as well, posing additional challenges.

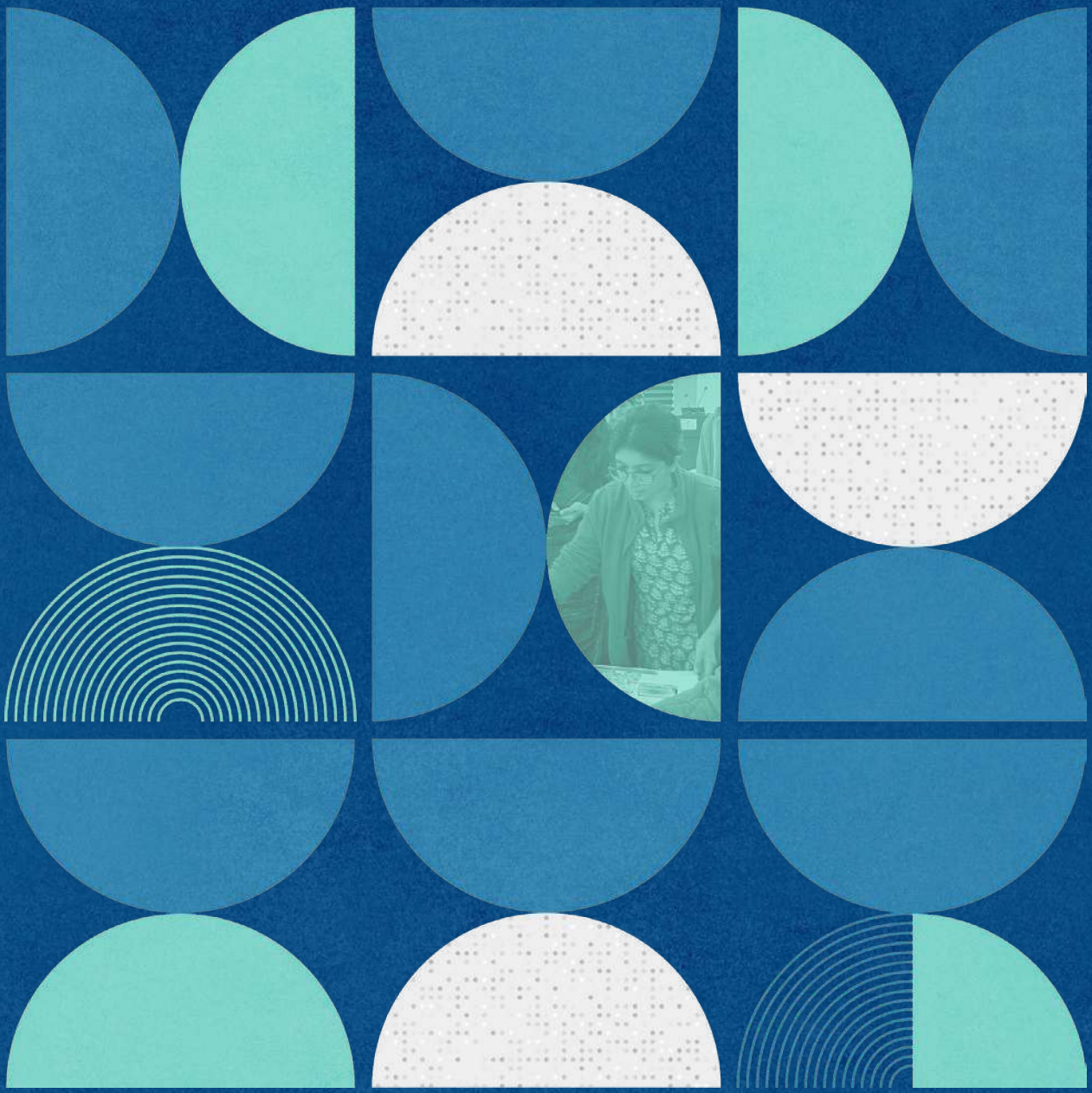
Building upon this intricate landscape of data ecosystems in Indian education, Mozilla and Aapti Institute embarked on the Strengthening Data Ecosystems (SDE) project. Funded by USAID, SDE is an initiative to enhance data stewardship practices in schools across India. At the heart of this initiative lay the dual objective of: promoting responsible data governance and safeguarding against the misuse or abuse of student data. In a landscape where educational data collection is becoming increasingly granular and widespread spanning various aspects, from attendance and test scores to discipline records, providing a detailed digital footprint of students' academic journeys. The potential of this data to transform educational outcomes is immense, yet it carries significant risks. Mismanagement, misuse, or inadequate storage of this data could lead to severe consequences, including surveillance, identity theft, or the perils of algorithmic bias. The SDE initiative aims to equip educators, parents, students, and administrators with the necessary tools to navigate, manage, and leverage this wealth of educational data effectively.

[20] Carlos Márquez-Vera, Alberto Cano, Cristobal Romero, Amin Yousef Mohammad Noaman, Habib Mousa Fardoun, Sebastian Ventura. "Early dropout prediction using data mining: a case study with high school students." November 2015.

[21] Data Information Knowledge Wisdom Pyramid

SECTION 2

Methodology



SECTION 2

Methodology

Recruitment of Working Groups

Our objective at Aapti Institute, in collaboration with the Mozilla Foundation, was to fortify the data ecosystems in Indian schools, and a crucial step in this endeavour was the recruitment of working groups consisting of academicians, education advocates and data scientists from public and private schools across India. The process we employed was thorough and deliberate, aimed at ensuring these groups were not just competent but also in sync with our overarching mission. The recruitment process began with a call for proposals, structured through a Google Form. This form served as the primary means for groups to express their interest and outline their qualifications.

Proposal Submission Process

Each working group, consisting of 3-4 members, was required to submit detailed information via the Google Form.

The information required included:

1. **Group Member Details:** Names of the working group members, their designations, and the organisations they were associated with.
2. **Location:** The city or state of residence of the members.
3. **Project Description:** A thorough outline of their proposed project, addressing specific points such as:
   - The goals of the project.
   - The issue areas the project aimed to tackle, focusing on either decision-maker fluency in data practices or understanding and objection of data use among those impacted by data-based decisions.
   - Strategies for achieving the project goals.

- Tangible outputs and deliverables, with a clear indication of how these could be utilised within the wider ecosystem.
- Methods for measuring the project's impact.
- Detailed descriptions of the core team, their relevant expertise, and the specific schools, communities, or students they intended to work with.

4. **Licensing Information:** The type of licence under which the project would be released, adhering to open-source criteria set by the Free Software Foundation (FSF) or the Open Source Initiative (OSI).

Evaluation Criteria

We received a total of 16 responses, representing 16 different groups comprising 86 individuals. These responses were evaluated based on a structured system with 10 distinct criteria, each with three sub-criteria. Each group could score a maximum of 30 points, with points allocated based on their compliance with each sub-point. Based on the evaluation, 5 working groups were selected for the initiative.

The criteria for evaluation were as follows:

1. **Mission Fit:** Assessed whether the proposed project advanced Aapti & Mozilla's mission to strengthen data ecosystems in Indian schools with possible focus on following areas: enhancing data literacy, evaluating the benefits and risks of data collection, empowering stakeholders to harness data in education, setting data governance standards, and creating tools for improved data access and understanding, supporting better decision making. Projects with no alignment scored lowest, while those with a clear advancement of the mission scored highest.
1. **Openness:** Evaluated the openness of the project, in terms of source code publication and licensing.
2. **Team Suitability:** Determined if the team was appropriately skilled and diverse (in areas like research, legal, design, coding) to execute the project.
3. **Connection with a School Community:** Checked for established connections and permissions with school communities.
4. **Potential to Scale:** Gauged the project's likelihood of growth over time.

5. **Impact:** Estimated the project's potential impact on data governance in education.
6. **Ability to Measure Impact:** Reviewed the clarity and adequacy of the plan to measure the project's impact.
7. **Ongoing Projects on Data/Education:** Considered existing projects of the team relevant to data, education, and data governance.
8. **Affiliations:** Assessed the working group's affiliations that could improve the project's quality and outreach.
9. **Diversity:** Looked at how the working group represented diverse perspectives.

Each of these criteria played a vital role in ensuring that the selected working groups were not only aligned with the initiative's goals but were also capable and equipped to make an impact in the realm of data governance in Indian schools. This process provided a robust foundation for the SDE initiative, ensuring its effectiveness.

The rationale behind this approach was multi-faceted:

1. **Addressing a Critical Need:** The rapidly increasing use of data in schools poses risks like surveillance, identity theft, and algorithmic bias. The SDE initiative aims to mitigate these risks and leverage data responsibly for educational advancement.
2. **Empowering Stakeholders:** By equipping educators, students, parents, and administrators with tools and knowledge, the initiative aims to transform data usage from a mechanism of oversight to a resource for empowerment.
3. **Cultivating Data Stewardship:** The focus on data fluency and responsible practices aligns with the larger goal of fostering a culture of data stewardship.

The expectations for the working groups were clearly outlined, focusing on three key areas: *Engagement with Local Communities,* requiring a commitment to actively work within specific school environments; *Time Commitment,* involving a dedication of 4-6 hours weekly, along with regular check-ins and mandatory attendance at a summit; and *Output Creation,* which entailed the development of tangible resources such as web applications, educational tools, and public engagement pieces. These outputs are designed to be useful to various stakeholders even after the

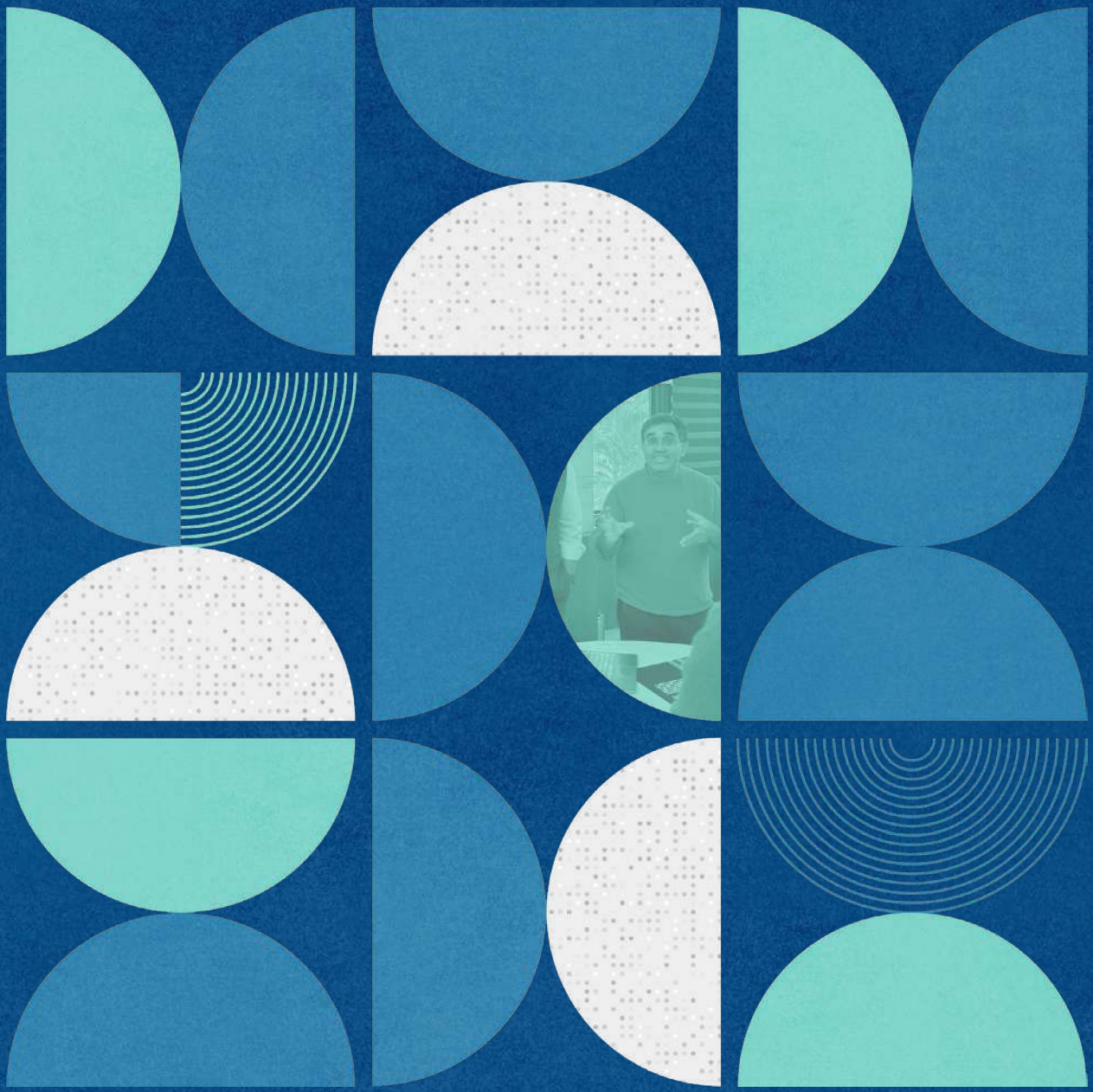
project's completion and the conclusion of the SDE workshop. This workshop, held in New Delhi from November 27, 2023 to December 1, 2023 marked the culmination of the working groups' collaboration with the Aapti Institute and Mozilla Foundation. To support the efforts of the working groups, a stipend of $2,500 USD was paid directly to each working group member, assisting with participation costs and travel expenses for the SDE workshop.

Managing the Working Groups

The five working groups selected were engaged from the month of August to November. There were bi-weekly meetings organised over Zoom, to understand the project progress, the objectives set by the groups, and the needs of the group members. October onwards, the check-ins were scheduled monthly. The focus of these meetings were to understand the readiness of the working group members for the SDE workshop.

SECTION 3

Project Profiles & Activities



SECTION 3

Project Profiles & Activities

Each working group, well-versed in India's educational ecosystem and current regulatory frameworks, initiated a project to explore the practical challenges associated with data in the sector. Following the evaluation process, five working groups were shortlisted. A brief overview and the objectives, of these five working groups along with their activities presented at the SDE Workshop in New Delhi has been provided below:

Working Group 1 (WG 1)

Project Title:

Towards Enhancing Data Utility and Accessibility for Schools

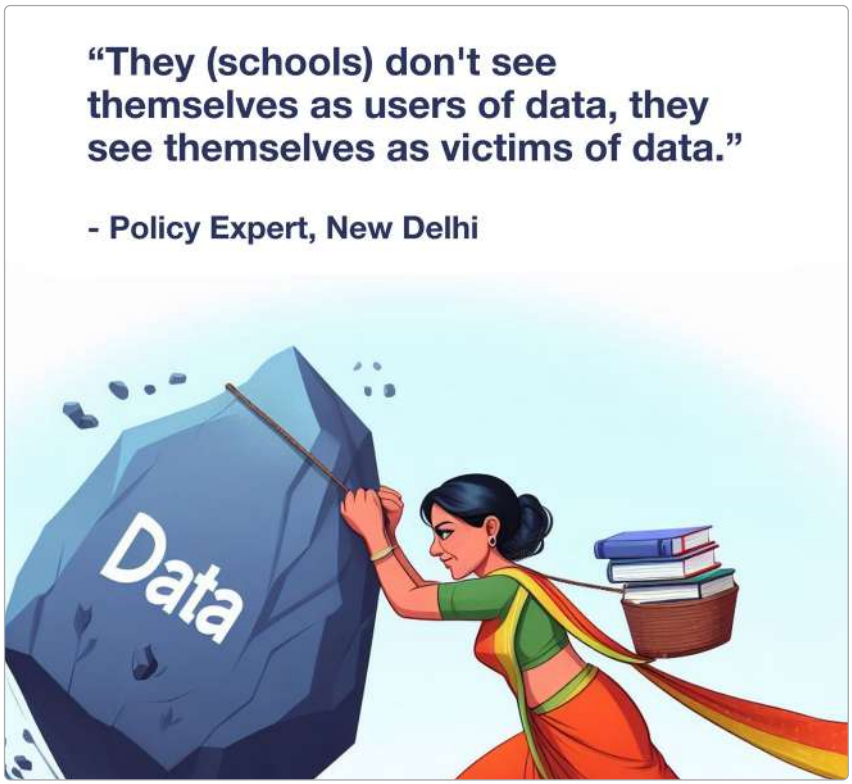


Image courtesy of Quicksand Design Studio.

Member Names, Profile and Objective

The core team of WG 1 included ***Nivedita Krishna, Tejasvini Vavilala,*** and ***Anagha Sasidharan***. Nivedita Krishna, serves as the Founder Director of Pacta, a Bengaluru-based boutique law firm that assists charities, foundations, NGOs, startups, and schools with their legal needs. Bringing over fifteen years of experience as a legal advisor in the public sector and non-profits, she focuses on law and policy initiatives to enhance India's public education system, in addition to her role as a technology policy consultant P. Tejasvini Vavilala, the Director of Projects at Sphoorti Foundation, is committed to helping underprivileged children through education and healthcare projects, leveraging her extensive decade-long experience in project management and strategic planning. Anagha Sasidharan brought with her experience in various technology policy initiatives, focusing on developing good governance models and legal frameworks for the just advancement of technology. The team also consulted ***Manasvini Ranganathan*** who has significantly contributed to the study through her expertise in interdisciplinary qualitative research. The team initially collaborated with a school operated by the Sphoorti Foundation in Hyderabad.

The objective of the WG 1 was to map out data mandated by the government from schools, compile databases such as the Unified District Information System for Education Plus [“UDISE+”], and applications for learning outcome measurements and attendance measurements. The goal was to serve students, shape better schooling outcomes, and move away from data as a means of oversight and punishment towards data for empowerment. Complementing this, they aimed to create a framework *‘Data Accessibility and Utility Framework (DAU)’* for assessing the accessibility and effectiveness of these government databases for the individuals responsible for compiling and contributing to the data.

Project Results

To achieve these objectives, the members studied school databases to maximise data potential, including for contributors. They conducted interviews with stakeholders involved in data collection to understand their experiences. The interviews were conducted in

schools located in the states of Karnataka, Tamil Nadu, Uttar Pradesh, Uttarakhand, Kerala and Assam. The group also reviewed consent protocols and data practices which effect a transparent and useful data ecosystem. They analysed the gaps to evaluate the potential of the data and how it can be used to empower schools.

Progress and Activities leading up to the SDE Workshop

MONTH	ACTIVITY
August	Established project timeline, reviewed UDISE data, explored NPSSE data sources. Prepared for interviews.
September	Completed majority of literature review, initiated interviews with a CBSE school in Uttarakhand and two practitioners. Developed data accessibility score.
October	Concluded interviews with five schools and ten practitioners, developed draft Data Utility Score Framework, engaged schools in Telangana and Andhra Pradesh.
November	Prepared working version of report, planned presentation for the SDE workshop.

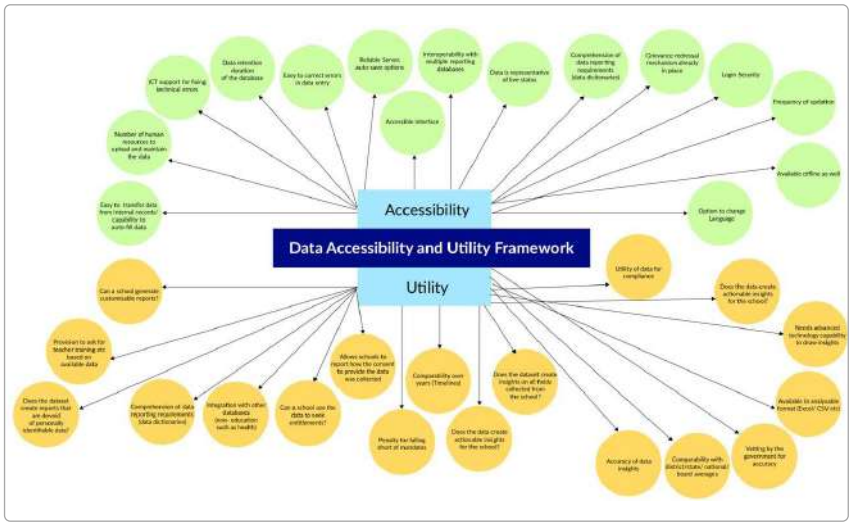
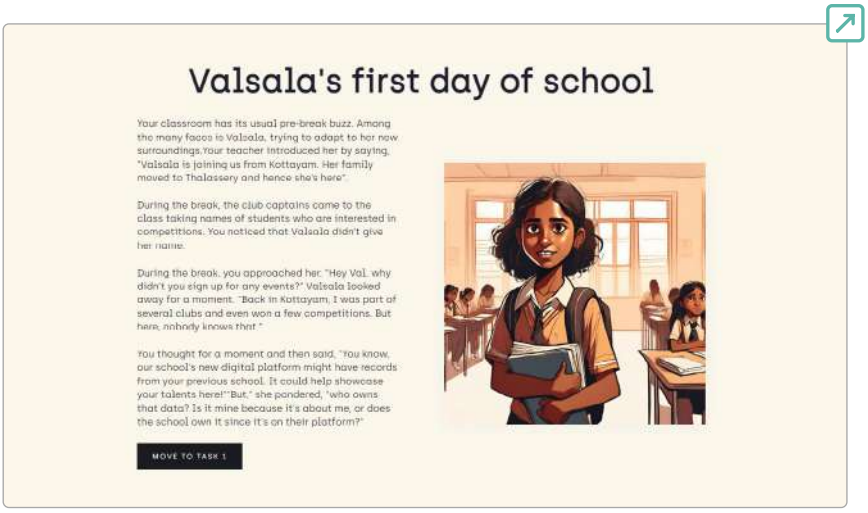


Image courtesy of Quicksand Design Studio and WG 1 members.

Working Group 2 (WG 2)

Project Title:

Enabling the development of localised understanding of data as a tool for empowerment in Indian schools through collaborative and story-based pedagogy



Screenshots from 'A Collaborative Learning Platform for Localised Understanding of Data' created by WG 2 members

Member Names, Profile and Objective

The Working Group 2 included ***Ajin K Thomas***, an experienced researcher, educator, and English-Malayalam children's book translator, who leveraged his expertise in various school contexts to facilitate the collaboration with local schools in Kerala. Aju Kurian, a senior product manager, brought valuable experience in

designing consumer and enterprise platforms, with his proficiency in user story development and wireframing critical for the project design and development. ***Suhair Killiyath Kuniyil***, a researcher in digital public goods and open data governance, mapped data flows and processes, contributing to the group's understanding of data governance in local contexts. ***Yesha Tshering Paul***, a lawyer and researcher specialising in digital rights, used her experience in public engagement and previous work on digital identity to ensure that the content was both impactful and accessible. Their combined expertise, particularly Ajin, Aju, and Suhair's proficiency in Malayalam, was pivotal for the group's output.

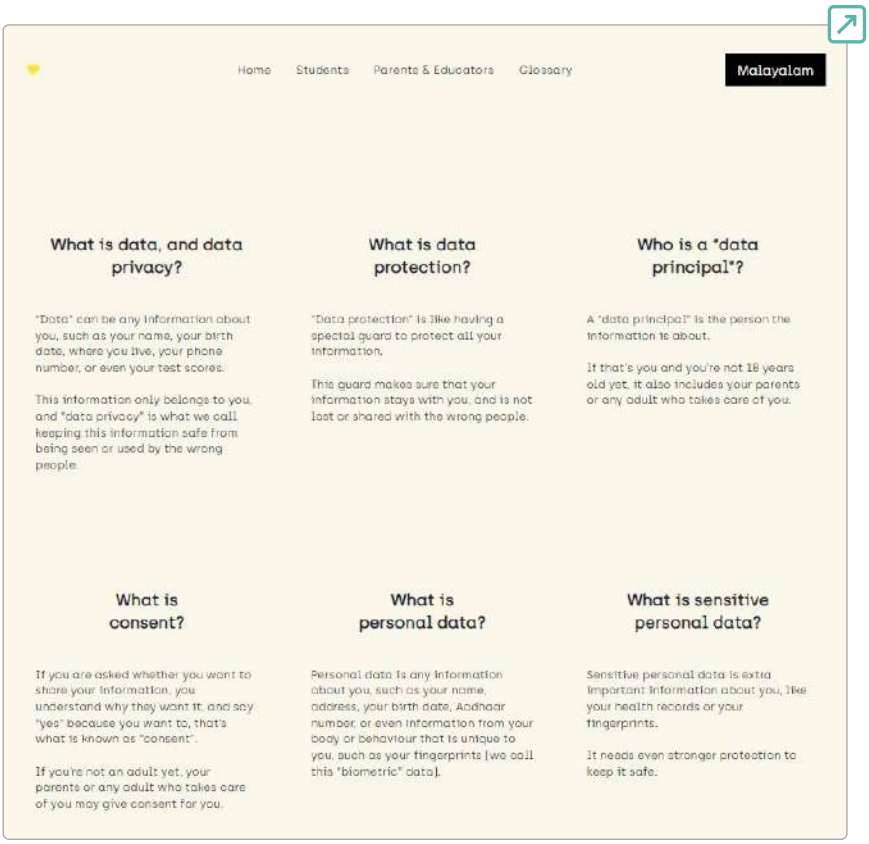
The objective of the WG 2 was to build a community-centric, collaborative platform, accessible in English and a vernacular language, to empower decision makers and data fiduciaries in advocating for better data practices. It focused on ensuring that data subjects in schools fully understand the risks and benefits of data collection and processing. The project targeted non-urban and vernacular educational contexts, responding to the lack of technological literacy in these areas. The scope was limited to in-depth consultations with stakeholders from a selected government-run school in Kerala, contributing to the research and development of the platform. This participatory approach intended to capture diverse stories and build a shared vernacular understanding of essential concepts through co-creation.

Project Results

The proposed output was a gamified learning platform promoting responsible data practices, featuring user personas, interactive decision-tree stories, a simplified data ecosystem map, a glossary of concepts in Malayalam, and a digital and print-friendly design. This platform was intended to build contextual understanding, empower critical examination of data and privacy, and deepen educators' understanding of these concepts. It was also designed to be independent of third-party mediation. The impact of the project was to be assessed through improved stakeholder understanding of data practices, using action research and direct school visits. This included evaluating the platform's influence on students' critical thinking and conducting interviews with teachers and parents to gauge its effectiveness and impact on existing data practices.

Progress and Activities leading up to the SDE Workshop

MONTH	ACTIVITY
August	Deliberated on research framework; began simplifying a glossary of terms on data protection and collection for kids.
September	Developed a basic website structure after interviewing teachers; analysed data collection platforms, planned school visits and workshops in Kerala; Completed the first iteration of the website design.
October	Shared the website prototype and began work on translating the content to Malayalam.
November	Finalised the translations while fine tuning the bilingual content on the website.



A screenshot of the Glossary from ‘‘A Collaborative Learning Platform for Localised Understanding of Data’ created by WG 2 members’.

Working Group 3 (WG 3)

Project Title:

Stor(y)ing Data Inaccuracy: Detechnicising our imagination of data production

Visual narrative excerpt from 'Stor(y)ing Data Inaccuracy' created by WG 3 members.

Member Names, Profile and Objective

The WG 3 included ***Joyeeta Dey, Nidhi K,*** and ***Vrinda Bhatia.*** Joyeeta, a PhD candidate at the National Institute of Advanced Studies, Bengaluru. She came with experience in studying the digital transformation of education governance in India, spanning from the National Informatics Centre's establishment to the National Education Policy (2020). ***Nidhi***, a Programs Associate at YLAC with a background in sociology, anthropology, and philosophy, brought with her skills in qualitative research and experience in research-based writing and communications. Vrinda, a PhD candidate at Jawaharlal Nehru University, New Delhi. Her expertise in managing visual and narrative data helped her understand the importance of combining quantitative and qualitative data. They also contribute to Aaina (**aainaedu.in**), a digital initiative using comics to disseminate educational research on social media.

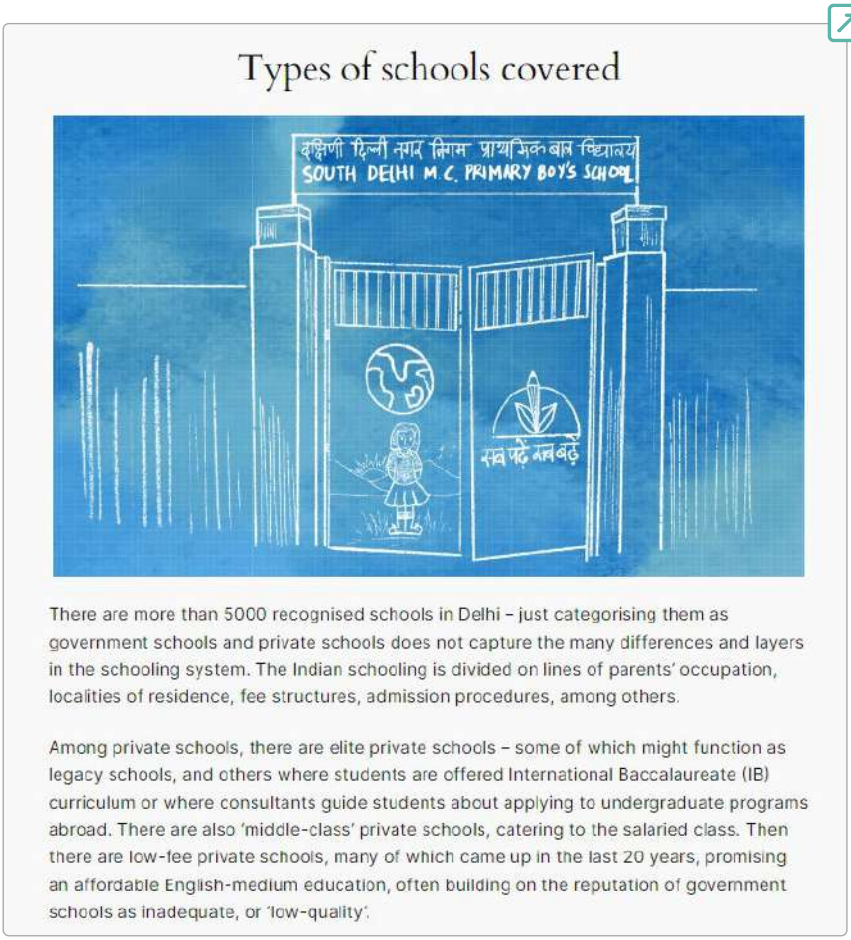
The objective of the WG 3 was to address the long-standing issue of data inaccuracies in education, traditionally seen as the result of either unintentional errors or intentional malpractice. They recognised that these inaccuracies stemmed from systemic issues rather than individual irresponsibility, and aimed to initiate a dialogue on the challenges of education governance data collection. This included addressing the burdens faced by teachers as data collectors, such as excessive non-teaching workloads, repetitive data requests, lack of input in data relevance, understaffing, insufficient infrastructure, and the pressure to manipulate data due to high stakes.

Project Results

The project employed visual storytelling to explore the social production of data inaccuracy and advocated for empathetic, participatory approaches to setting data collection agendas. By using comics to stimulate focus group discussions with school functionaries, the team sought to gain insights into the data collection process. These insights were then shared through **visual narratives** to demystify the process of governance data production in schools. With the aim of making accurate data more meaningful and applicable, rather than an elusive ideal.

Progress and Activities leading up to the SDE Workshop

MONTH	ACTIVITY
August	Finalised research outline. Reached out to academics for preliminary interviews and built a reading list.
September	Commenced interviews with school principals and IT members from Delhi Government schools. Began recruitment for the first focused group discussion (FGD).
October	Concluded three FGDs, identified the themes for visual narratives and began the illustration work.
November	Conducted follow-up interviews for clarifications on few processes related to data collection and editing. Started work on developing the **website** and drafting the stories for illustrations.

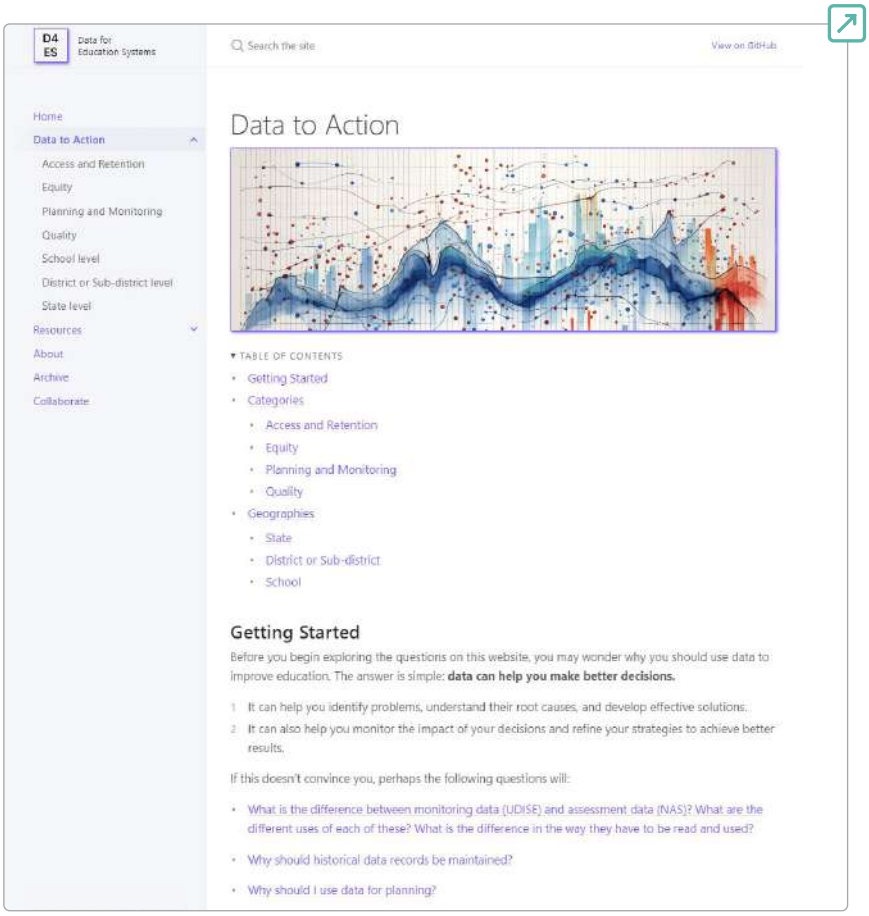


A screenshot showcasing the approach from the 'Stor(y)ing Data Inaccuracy' website developed by WG3 members.

Working Group 4 (WG 4)

Project Title:

Data for Education Systems in India: Design, Processes, and Use



A screenshot from *'Data for Education Systems'*, an interactive website created by WG 4 members

Member Names, Profile and Objective

The core team of WG 4 was led by ***Subir Shukla,*** accompanied by ***Deepti Srivastava, Gurjot Sidhu,*** and ***Ritwik Shukla.*** Subir, previously serving as the Educational Quality Advisor to the Ministry of Education, Government of India, utilised his extensive experience in engaging with national and state governments to implement large-scale educational quality improvement initiatives across India. Deepti came with experience in designing and facilitating large-scale assessments and research, contributing her substantial knowledge in this area. Gurjot has in the past supported the design and development of state data systems and

building the capacity of those involved. Ritwik came with research experience on health and education having extensively worked with state governments to enable policy- and decision-makers to use their data better.

For developing the guide, the members engaged with: Early Childhood Care and Education centres (Anganwadis) co-located in schools, teachers, head teachers and community representatives (SMCs) of schools, block and district functionaries—both administrative (BEO, District Education Officer) and academic (BRC, District Institute of Education and Training) and state level functionaries involved with educational planning and decision-making. They combined the research with their ongoing work in the states of Uttar Pradesh (ECE development), Rajasthan (school education curriculum and teacher development), Goa (system-wide assessment) and Andhra Pradesh (training of functionaries in Annual Work Plan and Budget Development).

The objective of WG 4 was to address the evolving data systems in India's public education. Previously, these systems, at Union, state, and local levels, were mainly used for accounting purposes, such as tracking entitlements and school resources. However, the challenge shifted towards using data to address quality concerns, especially after the disruption caused by the COVID-19 pandemic. The project's aim was to create a digital guide for government officials, policymakers, and researchers. This guide would assist in designing efficient data systems and using data to improve educational quality, including implementer performance and learning levels. The group planned to develop an interactive website containing principles and practical examples for designing context-specific data systems in education.

Project Results

The output was an **interactive website featuring learning tools** designed to elucidate the nature and objectives of an education data system. These tools included principles for system conceptualization, methods to contextualise goals and requirements, various data system examples, the roles of stakeholders in data utilisation, and illustrations of how data can enhance education quality. Insights from government officers,

education researchers, and teachers were used to validate these tools.

The impact of these tools was envisioned to serve various stakeholders, from high-level policymakers to teachers. The impact was to be measured by the website's viewership and stakeholder feedback.

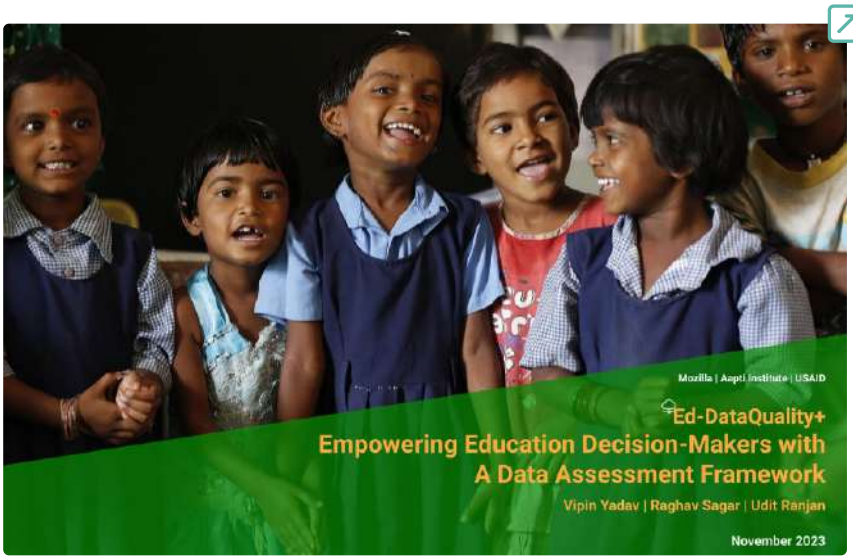
Progress and Activities leading up to the SDE Workshop

MONTH	ACTIVITY
August	Established project timeline, commenced secondary research, focusing on content development and building a library of case studies.
September	Created a matrix for key education data collection areas, identified stakeholders, circulated a survey with 80-100 questions, observed a gap in data-mindedness, and began work on website design.
October	Made progress in developing the website prototype and answers to the questions they received from the end users.
November	Completed and edited answers focusing on data collection principles, issues, and solutions. Ready to take the website live.

Working Group 5 (WG 5)

Project Title:

Ed-DataQuality+: Empowering Education Decision-Makers with data quality assessment Framework



Presentation on Ed-DataQuality+ by WG 5 members.

Member Names, Profile and Objective

The members of the WG 5 were ***Raghav Sagar, Udit Ranjan,*** and ***Vipin Yadav,*** each bringing a wealth of experience and expertise in education systems and interventions. Raghav, with a bachelor's degree in economics from the University of London, comes with nine years of experience in education interventions in India, specialising in creating data-based monitoring systems and implementation. He is part of the Classroom Instructions and Practice Team at the Central Square Foundation, India. Udit, a London School of Economics graduate in Economics and Philosophy, too has over nine years of experience in evaluating social programs in South Asia, with a focus on impact evaluation and assessment. He works with the Evaluation team at the Central Square Foundation. Vipin, holding a master's in Public Policy from the National University of Singapore, contributes over eight years of experience in education systems and interventions across various countries, with a specialisation in education intervention and evaluation. Through their project, they aimed to collaborate with stakeholders at the block level and selected schools in either Delhi

or Uttar Pradesh, focusing on local contexts to improve data practices and decision-making in education.

Through this project, WG 5 members aimed to address inefficiencies in existing frameworks for assessing education program data quality in India. These frameworks often lack contextualisation and compatibility with local data systems.

Project Results

The goals were to develop a contextualised and standardised framework for measuring the quality and usage of school education data at the block/district level and to implement this framework in a pilot study in a selected block/district in Delhi or Uttar Pradesh. The overarching aim was to equip decision-makers with the necessary knowledge and skills to advocate for data practices beneficial to students.

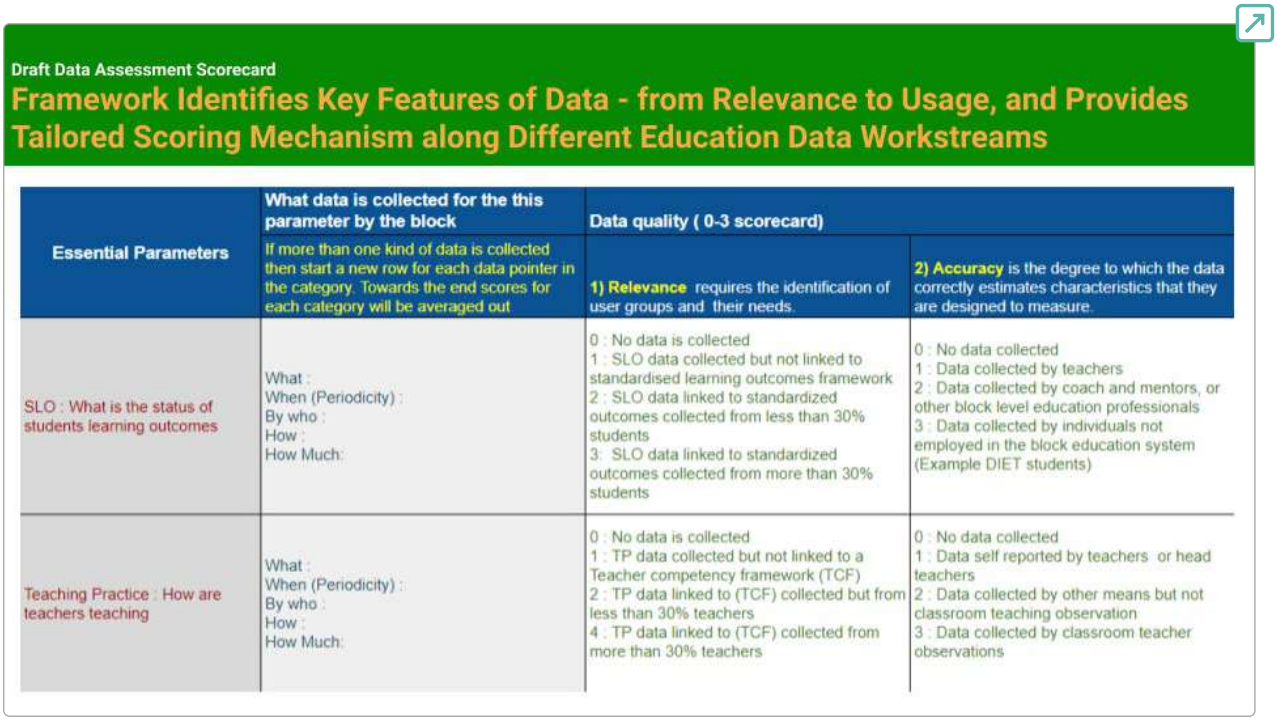
They worked towards three key outputs:

1. **A Contextualised Framework for education data quality assessment:** A rubric-based system that rated education monitoring systems on various criteria.
2. **A Study Report from the framework pilot:** This included field research findings and a comprehensive report with actionable recommendations.
3. **A "Bluebook":** A best practice document detailing the framework, its usage, and tools for implementing a data monitoring system.

The members intended for the framework to be used by block and district education leaders for informed decision-making. The impact was to be measured across three indicators: *accurate data quality assessment, ease of usability by block-level stakeholders,* and *usefulness of recommendations* compared to other established benchmarks. Ultimately, the bluebook has been planned to be made publicly available for broader stakeholder use.

Progress and Activities leading up to the SDE Workshop

MONTH	ACTIVITY
August	Initiated secondary research and began selecting a school community for engagement.
September	Noted the lack of streamlined data usage within education systems and the complexity of existing frameworks from UNESCO, World Bank, Harvard, and Oxford. Started developing a simplified framework and created a project plan with ongoing secondary research.
October	Conducted field visits for qualitative research in UP, with secondary research aiming for completion by mid-October to begin expert interviews.
November	Finalised the presentation and the rubric, incorporating feedback from expert discussion. Working to further refine the tools and research while testing the framework in selected blocks in Uttar Pradesh (UP) and Madhya Pradesh (MP).



Draft Data Assessment Scorecard

Framework Identifies Key Features of Data - from Relevance to Usage, and Provides Tailored Scoring Mechanism along Different Education Data Workstreams

Essential Parameters	What data is collected for the this parameter by the block If more than one kind of data is collected then start a new row for each data pointer in the category. Towards the end scores for each category will be averaged out	Data quality (0-3 scorecard) **1) Relevance** requires the identification of user groups and their needs.	 **2) Accuracy** is the degree to which the data correctly estimates characteristics that they are designed to measure.
SLO : What is the status of students learning outcomes	What : When (Periodicity) : By who : How : How Much:	0 : No data is collected 1 : SLO data collected but not linked to standardised learning outcomes framework 2 : SLO data linked to standardized outcomes collected from less than 30% students 3 : SLO data linked to standardized outcomes collected from more than 30% students	0 : No data collected 1 : Data collected by teachers 2 : Data collected by coach and mentors, or other block level education professionals 3 : Data collected by individuals not employed in the block education system (Example DIET students)
Teaching Practice : How are teachers teaching	What : When (Periodicity) : By who : How : How Much:	0 : No data is collected 1 : TP data collected but not linked to a Teacher competency framework (TCF) 2 : TP data linked to (TCF) collected but from less than 30% teachers 4 : TP data linked to (TCF) collected from more than 30% teachers	0 : No data collected 1 : Data self reported by teachers or head teachers 2 : Data collected by other means but not classroom teaching observation 3 : Data collected by classroom teacher observations

Screenshot from the presentation on Ed-DataQuality+ by WG 5.

Key Insights Shared by the Working Groups

Historical Challenges in Data Utilisation

The discussions provided historical context on data and data collection, highlighting a period between 1995 and 2003 when there was a rush to report educational data, motivated by funding requirements. This led to inaccuracies and a reactive approach to data collection and utilisation. This early phase was characterised by pressure on administrative bodies to hastily report data, and set a precedent for the perception and utility of data within the education sector. It was seen as a bureaucratic requirement rather than a tool to enable effective decision making and improvement at the local level.

Evolution of Data Use

Over time, the parameters for funding evolved to include learning outcomes, signalling a shift towards recognising the importance of data in improving educational quality. This shift also brought about an increase in the complexity of tools and the volume of data, leading to a fragmented landscape with various datasets, often overlapping (e.g., NAS, UDISE, PGI). This complicated the interpretation and effective use of data.

Utility of Data for Budgeting

There exists, even now, a significant gap between the collection of education data and its practical utility for schools, particularly in budgeting and operational planning. Despite collecting vast amounts of data, schools lack the readiness and capacity to analyse and use this information effectively.

Challenges to Data Collection and Utilisation

Schools face multiple obstacles in data collection and utilisation, including limited internet access, reliance on manual data entry, frequent system crashes, and a cumbersome data entry process. Additionally, there is a lack of trust in the data collected, further worsened by inadequate training for teachers and administrative staff.

Integration of Education Data with Other Databases

There exists a potential for enhancing the utility of educational data by integrating it with other databases, such as health data. For instance, UDISE data, when combined with health or socioeconomic data, can provide granular insights into the challenges and needs of students, particularly in conflict zones or underprivileged areas. Such an integration can help develop targeted and effective educational strategies and resource allocation. This insight highlighted the importance of interoperability between different data sets and the need for a multidisciplinary approach to educational planning and data analysis. Schools and NGOs that group members interacted with, expressed the need for detailed comparative data, which can be used for improved educational outcomes.

Localised Approach to Data Education

The presentation on the collaborative learning platform for localised data understanding in Kerala was an approach that prioritised local contexts and collaborative learning to democratise data literacy. By developing a platform in the vernacular language-Malayalam, the group aimed to make complex data concepts accessible to a broader audience, fostering an inclusive discourse around data in schools.

Challenges to Data Governance

There exist challenges in data governance and literacy within schools, such as prevalent gatekeeping, hierarchical structures, and an overarching emphasis on data collection over pedagogical priorities. Teachers and IT administrators often function merely as data collectors, with a limited grasp of the principles related to data minimisation[22] and access controls. Highlighting the need for responsible data governance frameworks within schools.

[22] Section 5,6 7 and 8 Digital Personal Data Protection Act, 2023

Human Aspect of Data Management

Through real-life scenarios depicted in comics, the presentation highlighted the importance of acknowledging the human factors involved in data management. Showing how teachers and administrative staff grapple with incomplete, changing, or inaccurate student data. The intent behind this approach was to humanise data workers, shedding light on the subjective experiences and challenges they face, which are often overlooked in discussions about data accuracy. By using visual narratives to explore the themes of data inaccuracy, the team aimed to foster empathy and understanding around the day-to-day processes involved in data management. This approach emphasised the need for discussions around data to incorporate empathy and a recognition of the emotional and practical realities faced by those responsible for managing educational data.

Causes of Data Inaccuracy

The research identified eight core themes contributing to data inaccuracy, including human behaviour, infrastructural failures, issues with data formats and definitions, overburdening of staff, bad planning, inadequate verification processes, and the challenges of centralisation in data management. These themes illustrated the multifaceted nature of data inaccuracies, suggesting that solutions must address a broader spectrum of factors than just malpractice or lack of training.

Limitations in Data Representation

A significant insight was the critical examination of language and numerical data as representations of reality. The observation that *"language is never a complete representation of reality, and numbers are one step short of language"* pointed to the inherent limitations in how data is collected, defined, and interpreted.

Absence of Data Mindedness

A critical insight was the concept of 'data-mindedness' – the ability and willingness to engage with data constructively. The absence of data-mindedness across various levels of the education system, coupled with a systemic inability to effectively use data for planning and improvement, was highlighted as a significant barrier. This issue is exacerbated by a general lack of understanding and

motivation to leverage data, despite its potential to inform better educational strategies.

Empowering Block Level Officers

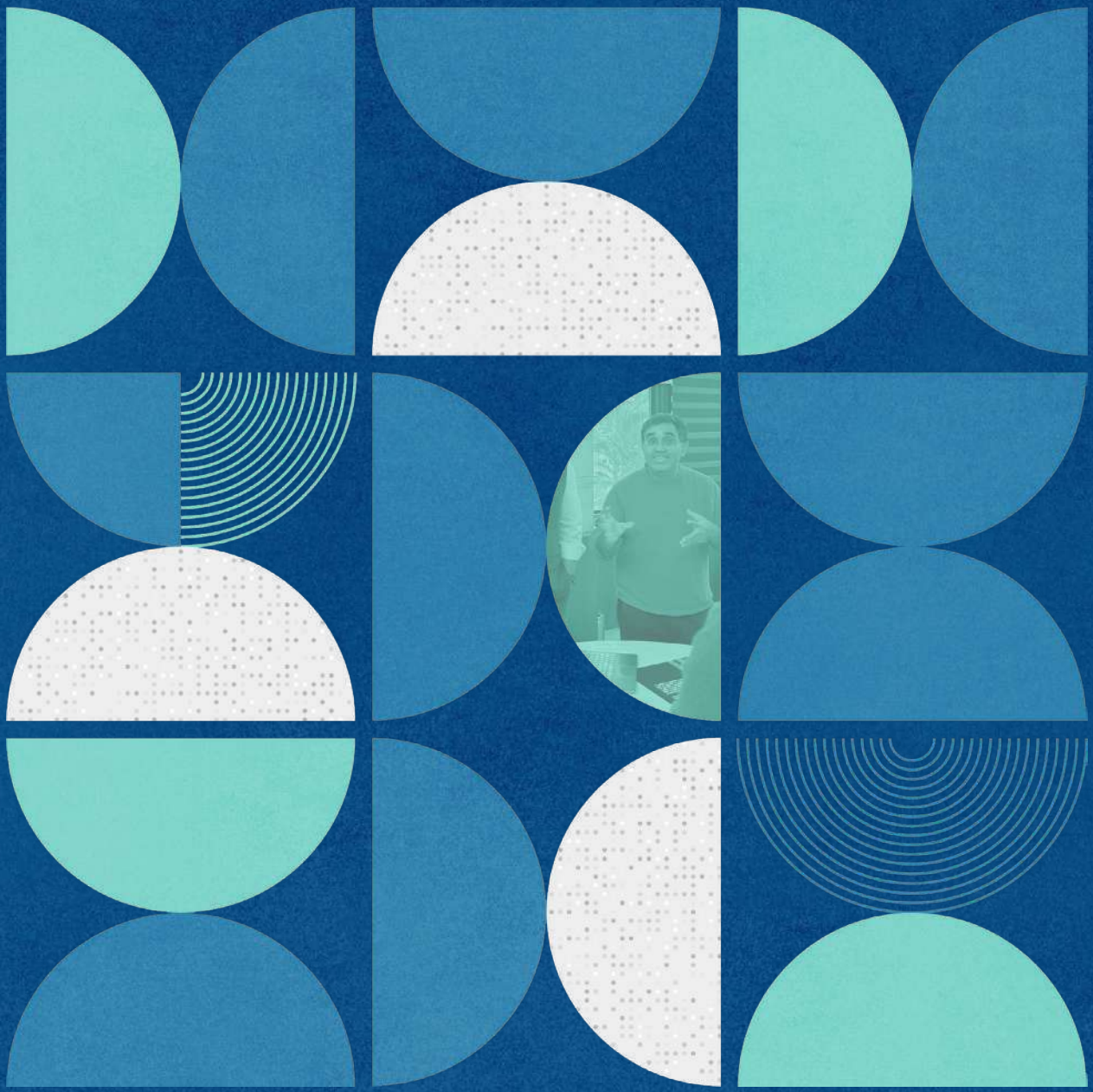
Another insight shared was that block level leaders often act merely as conduits for transferring orders from district officials to teachers, without the autonomy to make decisions that could lead to improved educational outcomes in their specific geographies. This lack of empowerment prevents the contextualisation of decisions, highlighting a systemic issue where data collection overshadows the strategic use of data for planning and improvement. By examining this issue, the necessity of empowering block level leaders to utilise data constructively was highlighted.

> **CSMS [the Centralized School Management System software] enables the teacher to get "everything" about a certain student. Including the history of the child (academic, psychological, medical)."**
>
> **- Teacher, Private School**

Image courtesy of Quicksand Design Studio

SECTION 4

Overview of the SDE Event



SECTION 4

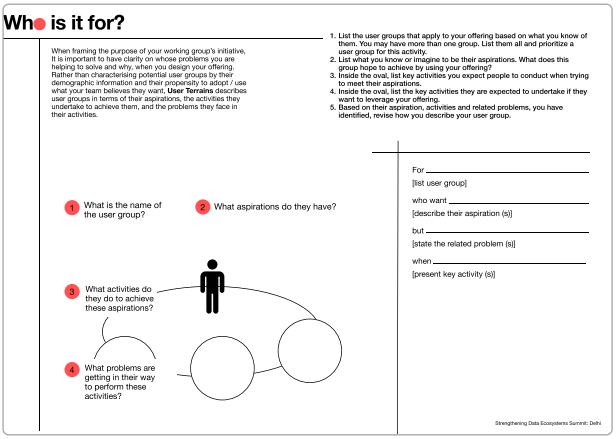
Overview of the SDE Event

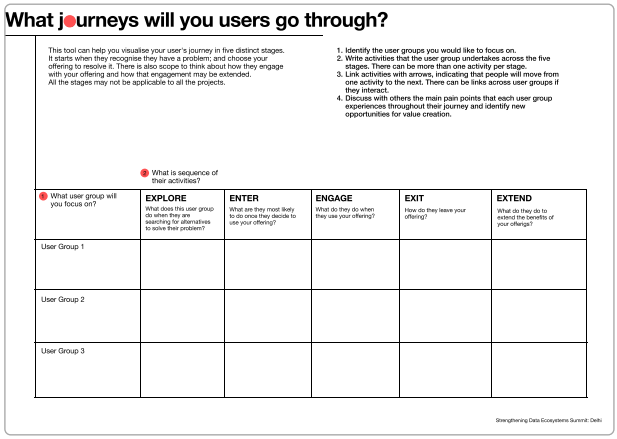
Culminating SDE Event

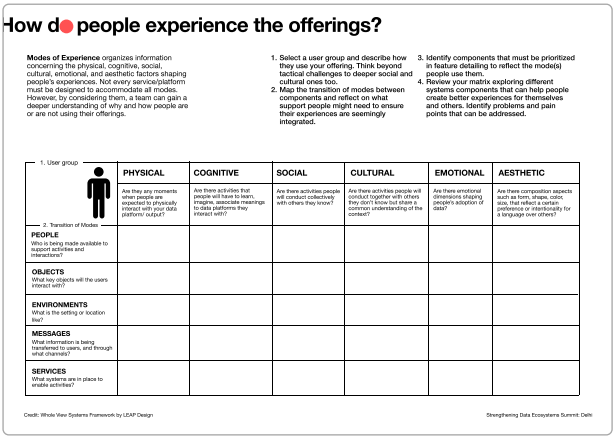
The Strengthening Data Ecosystems in Indian Schools workshop, was held from November 27 to 30, 2023, at the India Habitat Centre in New Delhi. The focus of the event was to provide a collaborative platform to the working groups. The groups shared their insights and discussed their learnings about the prevailing data culture in Indian schools. The agenda of the event, designed in collaboration with Quicksand Design Studio, can be found in the annexures *(Annexure 1 and 2)*.

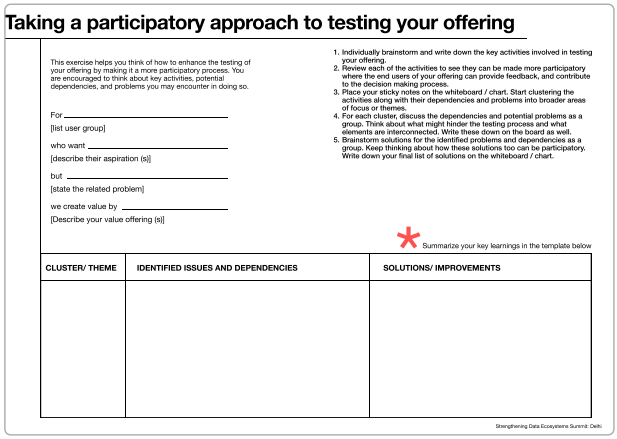
Left to right: Astha Kapoor, *Co-founder at Aapti Institute*; Roshan Nair, *Digital Inclusion Specialist, USAID*; Dr. Ganesh Kumar Nigam, *Education Specialist, UNICEF*; Karthik Menon, *Lead- Government Partnerships, Pratham Education Foundation*; Archana Sivasubramanian, *Lead- Centre for Technology Law and Policy, Sai University* during the Panel Discussion.

Quicksand leveraged their expertise in user-centred innovation to develop design tools for the workshop, encouraging critical thinking leading to an interactive discussion. These tools enabled working groups to gain a deeper understanding of their target users by exploring their aspirations, activities and challenges, looking past basic demographics for more insightful user needs. They mapped user journeys across various stages, from problem recognition to extended solution engagement, thus enhancing both design and engagement. This method encouraged working groups to think about aligning their project outputs with user expectations and effectively meet user needs. Moreover, the design tools focused on a participatory user engagement through structured feedback, and closely aligning solutions with user needs by identifying key activities and dependencies.









Tools for design exercises, created by Quicksand Design Studio.

Key Insights Shared During the Panel Discussion ‘Unlocking the Value of Education Data for Societal Good’

Panellists

1. **Roshan Nair,** Digital Inclusion Specialist, USAID
2. **Dr. Ganesh Kumar Nigam,** Education Specialist, UNICEF
3. **Karthik Menon,** Lead- Government Partnerships, Pratham Education Foundation
4. **Archana Sivasubramanian,** Lead- Centre for Technology Law and Policy, Sai University

The panellists highlighted key aspects of education data management in India. While India has comprehensive data from primary to secondary levels, which aids in understanding education trends such as access, retention, and quality, there are challenges like limited preschool data and over-reliance on pdf reports, which make data less accessible. The panel discussed the need for improved data analysis and data visualisation tools, while acknowledging that over collection of data without proper analysis or utilisation could be counterproductive. The discussion stressed the importance of empowering end-users such as teachers to understand and use the data collected effectively. Recognising these overarching themes, the panel delved deeper into specific areas of improvement and opportunities:

Inaccessibility and Utilisation

A significant amount of data exists in non-user-friendly PDF formats, making it challenging for researchers and stakeholders to analyse and derive clear insights. There is a need to transition from traditional formats to modernising data presentation. Making interactive and actionable. This can be done by leveraging data analytics tools to transform raw data into insights.

Overcollection vs. Utilisation of Data

The current practice of over collecting data without clear utility was identified as problematic. There is a need for rationalising data collection efforts to focus on meaningful metrics that directly inform educational strategies and outcomes.

Incorporating Technology for Data Visualisation and Analysis

The potential of technology to make data more accessible and understandable is acknowledged but underutilised. There is a significant opportunity to employ data visualisation tools and analytics to present data in formats that are more easily digestible and actionable for a wide range of stakeholders.

Building Data Literacy

There is an emphasised need for a comprehensive revision of the teacher education curriculum to include data literacy. This is crucial to prepare educators for a holistic assessment approach, which extends beyond traditional academic metrics to include communication, behaviour, extracurricular activities, and emotional well-being. Equipping teachers with the skills to interpret and apply this diverse range of data will enable them to inform their teaching practices and support the overall development of students, aligning with the broader objectives of the NEP.

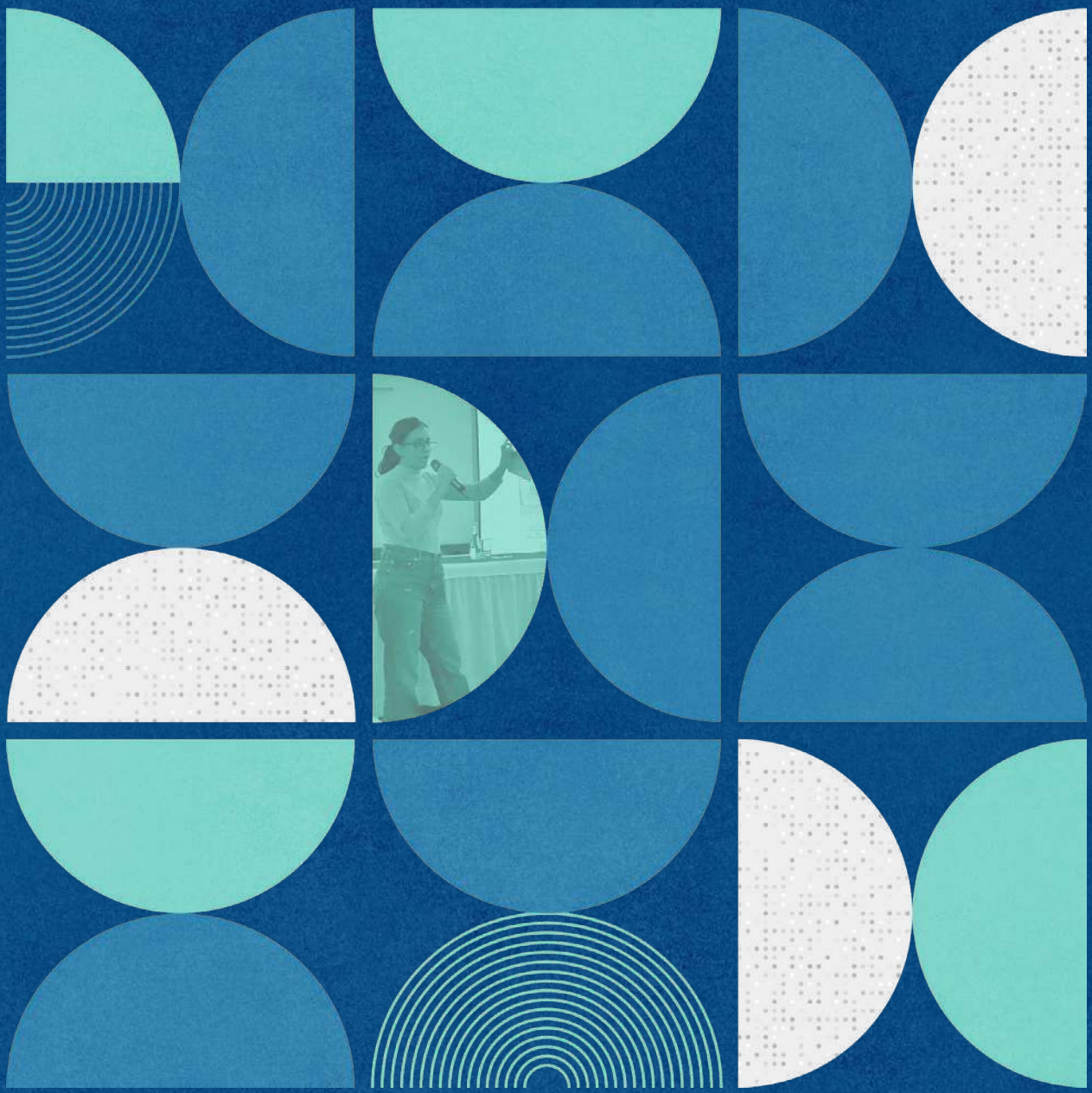
Community and Stakeholder Engagement

Increasing involvement of communities and stakeholders in understanding and using data. The discussion also highlighted the necessity of third-party verification to enhance the quality of educational data. Assessment and validation of data by independent means can ensure data accuracy. This process not only secures data integrity but also encourages community involvement, and can be a crucial step towards making a trustworthy foundation for informed policymaking.

The session was moderated by **Astha Kapoor,** Co-founder at Aapti Institute.

SECTION 5

Challenges



SECTION 5

Challenges identified through the SDE initiative

The SDE initiative provided us at the Aapti Institute the opportunity to engage with the working group members as they developed their project outputs. Through the event, we had the opportunity of interacting with the panellists and experts in the field of school education data. Additionally, our research into India's existing data system has allowed us to identify several challenges related to the collection, management, and even the basic understanding of school education data. To provide readers with context, we have included in this section a brief analysis along with a list of current projects (indicative, not exhaustive) and datasets pertinent to the school education data landscape in India.

Current State Analysis

Under the ambit of NEP 2020, NDEAR's vision is to revolutionise the Indian education system through a unified and interoperable digital infrastructure. This transformative approach is not only limited to digitisation but is also aimed at creating a synergistic ecosystem that will bring together disparate data points into a cohesive and actionable framework. Introduced by the Central Government as part of the National Education Policy, NDEAR is conceptualised as an architectural blueprint for the educational ecosystem, rather than a singular platform or application. It outlines a set of principles, standards, specifications, and guidelines—focusing on aspects such as technology, data process, and ecosystem engagement, with an emphasis on data, openness, inclusion, and accessibility. It aims to identify the essential building blocks that are critical for realising this blueprint into a tangible reality. NDEAR facilitates an ecosystem where various actors can develop, and innovate solutions, platforms, and tools that are

[22] Section 5,6 7 and 8 Digital Personal Data Protection Act, 2023

interoperable. The framework enables these solutions to advance policy goals through educational programs. NDEAR's framework is built on 12 guiding principles aimed at enhancing the education ecosystem's digital solutions with a focus on interoperability, transparency, scalability, and security. These principles include adopting the India Enterprise Architecture framework, employing open-source and open standards for services, ensuring portability of data and credentials, and maintaining scalability across automated and distributed processes. Resilience through automated recovery, leveraging emerging technologies, ensuring open data availability for research, utilising minimal and reusable microservices, and upholding security, privacy, and trust. Moreover, NDEAR stresses having decentralised registries and master codes for interoperability, and designs for solutions that are evolvable. These principles emphasise that NDEAR is a comprehensive blueprint for an interoperable, transparent, and secure educational architecture that encourages innovation while being adaptable and user-centric.

Considering the phased rollout and the ongoing early-stage implementation, it would be premature to delve into a detailed analysis of the framework. However, the challenges to its implementation such as:

1. Sub-optimal digital infrastructure and capacity in schools and rural areas
2. Risk of non-uniform adoption across states, as education is a state subject
3. Limited time and capacity for teachers for digital learning, as they are overburdened with administrative tasks have been included in other sections of the report.

Strategic Sub-Projects

1. **Vidya Samiksha Kendras (VSKs):** Central to NDEAR, VSKs facilitate the integration of state and national data systems. Acknowledging that the existing data sets such as- UDISE, NAS, NIPUN Bharat, Teacher data bases, and DIKSHA operate in silos, through VSKs there is an attempt to enable an integrated framework which allows - data sharing, enabling different data sets to speak to each other, leveraging them for

data-based decision making to eventually improve learning outcomes[23]. VSKs are being established at state level, to track student enrollment data and learning outcomes from to analyse retention and dropout rates, track academic and non-academic activities at state level. The initiative will also monitor the real- time status of existing schemes under the ambit of Samagra Shiksha. VSK also provides for digitisation of data from physical documents like attendance register, mark sheets etc to shift from the culture of manual data collection and data entry to digitisation of records, the initiative also seeks to establish a centralised dashboard displaying real-time performance indicators of schools. Through these steps issues that need urgent attention at school and state level shall be identified. Distribution of incentives through various welfare schemes like scholarships, uniform and free textbooks will also be tracked through VSK. VSK shall also include data from the PM-POSHAN programme. The NITI Aayog's report on best practices in the social sector highlighting innovative models across India, includes the implementation of VSK in Gujarat as a best practice. VSK in Gujarat exemplifies the effective use of data analytics in monitoring educational outcomes. By analysing data from various sources, VSK provides insights into student performance, teacher effectiveness, and resource allocation. This model of data-driven decision-making in education can be pivotal in enhancing educational quality and governance. At the moment VSK has been implemented in 12 states (Andhra Pradesh, Arunachal Pradesh, Delhi, Gujarat, Goa, Himachal Pradesh, Maharashtra, Nagaland, Punjab, Odisha, Uttarakhand, and Uttar Pradesh)[24]. However, replicating this success on a national scale through NDEAR faces challenges including standardising data collection and ensuring uniformity in implementation across diverse educational landscapes. In addition to the challenge of standardising data formats across diverse state systems, there is a need to ensure consistency in data collection, and maintaining data integrity across platforms.

2. **Gati Shakti[25] and Geographical Information System (GIS) Mapping:** While the aim of Gati Shakti is to optimise resource allocation through geographical mapping by identifying infrastructural and logistical assets, the challenge lies in the

[23] Vidya Samiksha Kendra

[24] "Third National Conference of Chief Secretaries on Ease of Living | SCHOOLING: ACCESS & QUALITY." Ministry of Education, Government of India, December 2023. Page 2.

[25] PM Gati Shakti

[26] Press Information Bureau, Union Education Secretary urges incorporating the study of the PM Gati Shakti scheme in academic programs, June 2022

accurate representation of remote and underdeveloped regions, often underrepresented in digital mappings. To this end, GIS mapping contributes and helps to the planning and decision making process especially relating to resource allocation. It is not only a visual tool but also a technology which helps build infrastructure on top of the other information from various perspectives. This can be utilised to determine requirements for schools in a scientific and practical manner[26]. Integration of Gati Shakti and GIS mapping is being done to facilitate the planning of education facilities. Selected PM SHRI[27] schools are linked with the Gati Shakti Portal. The portal also supports Pradhan Mantri Janjati Adivasi Nyaya Maha Abhiyan (PMJANMAN) to improve education quality, enrolment, and retention among vulnerable tribal groups.

3. **Project Appraisal, Budgeting, Achievements, and Data Handling System (PRABANDH)[28]:** PRABANDH offers a framework for project assessment and budgeting, critical for efficient resource allocation. It is a step to boost efficiency and manage the implementation of a centrally sponsored scheme for schooling- Samagra Shiksha, through technology. PRABANDH aims to eliminate submission of hard copies, unless where expressly required.

   It allows for online submission of the following

   - Online submission and appraisal of workplan and budget,
   - Online generation of sanction order regarding funds released by the Government of India,
   - Online submission by the states of monthly progress reports on the educational initiative Samagra Shiksha, covering both physical and financial aspects, including fund transfers and budget management,
   - Online submission of progress reports detailing functional and construction statuses across multiple Samagra Shiksha components, accompanied by photographs.

4. **Samagra Shiksha Scheme[29]:** is an overarching programme for the school education sector extending from pre-school to class 12. It is a centrally sponsored scheme funded jointly by the central and state governments, following a shared funding

[27] PM SHRI Schools 'The PM SHRI School initiative, sponsored by the Government of India, aims to upgrade over 14,500 schools across the country, ensuring a resource-rich environment for students by ensuring good physical infrastructure and appropriate resources. Targeting more than 20 lakh students, it focuses on fostering inclusive growth and preparing engaged citizens in line with the National Education Policy 2020. This scheme, set to run from 2022 to 2027, also seeks to enhance the quality of education and provide insights for policy and practice improvements nationwide'

[28] Department of School Education and Literacy, PRABANDH

[29] Samagra Shiksha Scheme

pattern[30]. The scheme provides essential financial support, enabling the holistic development of education at the grassroots level. The vision of the Scheme is to ensure inclusive and equitable quality education from pre-school to senior secondary stage in accordance with the Sustainable Development Goal (SDG) for Education. It consolidates the three previous programs: Sarva Shiksha Abhiyan (SSA), Rashtriya Madhyamik Shiksha Abhiyan (RMSA), and Teacher Education (TE)[31].

5. **Child Tracking System:** The Ministry of Education, in collaboration with States/UTs and Autonomous Bodies, is developing registries for students, teachers, and schools to comprehensively monitor students, to generate data based interventions. To date, 29 States/UTs have implemented these registries online, 17 States/UTs have fully established student registries, five are currently in the process, and fourteen are in the development phase[32]. Introducing APAAR ID (Automated Permanent Academic Account Registry), a unique student identification system linked with Aadhaar[33], marks a move towards enhancing the monitoring and management of a student's educational trajectory. This system's integration with existing platforms like UDISE+, VSK, and the Academic Bank of Credits (ABC) aims to provide a comprehensive view of a student's academic journey, enabling better service delivery and learning outcomes by identifying and addressing dropout causes. This initiative is expected to result in effective educational governance. However, experts have also questioned the need to have a distinct APAAR ID when Adhaar has already been deemed mandatory for school enrolment. As the purpose and objectives sought to be achieved through APAAR ID can easily be achieved and have to a certain extent been achieved through Adhaar linkage.

[30] Samagra Shiksha Scheme

[31] Samagra Shiksha Scheme

[32] "Third National Conference of Chief Secretaries on Ease of Living | SCHOOLING: ACCESS & QUALITY." Ministry of Education, Government of India, December 2023.

[33] 'What is APAAR ID? How is different from Aadhaar Card?' LiveMint, Jan 2024

Brief Overview of Existing Central Data Sets along with a few State Data Sets[34]

UDISE+

- **Scope:** National (Central Government Initiative)
- **Description:** Comprehensive data collection system for school education in India, covering infrastructure, student achievement, teacher and staff details, and more. Crucial for planning, resource allocation, and policy implementation.

Delhi (Del E or Edudel)

Del E, or Edudel, is the digital platform for Delhi schools, facilitating information sharing and data management through the DoE app[35]. It serves students, teachers, and administrators by offering access to educational materials, attendance tracking, and student data management.

Maharashtra (SARAL)

SARAL focuses on real-time data collection, managing transfers, recording assessment results, and managing various educational services. Complements UDISE+ with its more frequent data updates.

Tamil Nadu (TN-EMIS)

TN-EMIS collects and manages educational data for the state, offering services related to the education system, including student and staff management and assessment results.

Rajasthan (Shala Darpan)

Shala Darpan provides a platform for managing school data, focusing on the transparency and efficiency of school administration in the state.

West Bengal (Banglar Shiksha)

An EMIS platform specific to West Bengal, focusing on managing educational data within the state.

[34] This list is provided to offer context to the readers and is not exhaustive.

[35] DoE app FAQ

Kerala Infrastructure and Technology for Education (KITE)

- Focuses on enhancing the infrastructure and technology use in education in Kerala. Includes managing digital resources and infrastructure data.
- SAMETHAM[36] The Kerala School Data bank, a KITE initiative, provides verified school data and is now live for schools to confirm their details, such as- approved classes, details of students in each class, teachers, non-teaching staff (excluding personal information).
- SAMPOORNA ERP an open source school management system developed.

Karnataka (SATs)

SATs is used for managing educational data, aiding in effective administration and planning.

Haryana (NIPUN)

Focuses on foundational literacy and numeracy, contributing data to inform educational strategies and policies.

Assam (Shiksha Setu Ahom)

Description: A local data management system, it is currently not synchronised with UDISE+, indicating potential gaps or delays in data integration.

Private Organisations

Annual Status of Education Report (ASER)[37] grounded in evidence-based decision making, ASER is a survey conducted by Pratham, it surveys the status of children's learning outcomes and enrollment in rural India.

[36] "Third National Conference of Chief Secretaries on Ease of Living | SCHOOLING: ACCESS & QUALITY." Ministry of Education, Government of India, December 2023. Page 27.

[37] Annual Status of Education Report, Pratham

Challenges

1. **Transitioning to a Student-Centric Ecosystem:** Shifting from a school-based to a student-based system demands a fundamental change in data architecture. Despite the initiatives announced and steps implemented under the NEP, 2020 to give effect to this transition, there is a need to address challenges in technical integration in how educational data is collected, managed, and utilised.

   - Data collection efforts at present are fragmented with multiple bodies and authorities reaching out to schools at different times in a year requesting for data; in many instances these multiple requests for data are for the same data set. This multiplicity of demand puts a burden on schools, which results in the data not being provided in a timely manner. Additionally, states view the central level data with distrust. This results in lack of ownership of data, where the providing data is looked at as a mere job with no regard for data accuracy. This highlights the inefficiency in the system.

   - A significant challenge is the mismatch between different datasets and the lack of real-time data. The data that exists is often outdated, which can lead to ineffective policy-making and resource allocation. These discrepancies undermine the reliability of educational insights and necessitate a streamlined approach to ensure data relevance and accuracy across various platforms.

2. **Cultural Shift Challenge:** Overcoming the problem of Data Rich and Information Poor (DRIP)[38] and building trust between teachers and administrators regarding the collection, handling and utilisation of data. Schools often struggle with the practical application of data, lacking a strong data culture. This is compounded by a limited understanding of data's potential beyond mere numbers, highlighting the need for a shift towards a more data-informed mindset.

3. **Barriers to Integrating UDISE+ with NDEAR's approach:** UDISE+ currently represents the most comprehensive database in the Indian education sector. Integrating it with NDEAR's framework necessitates not only technological upgrades but

[38] Data Culture Guide- Edelements

also a rethinking of data utilisation strategies. The digital divide poses a significant barrier, especially in rural and underprivileged regions, with poor ICT facilities potentially leading to data gaps which result in skewed policy decisions. Schools also face technical barriers in efficiently collecting and reporting data, hindering effective data management practices. It is also important to note that most government schools are located in regions where teachers, who are also in charge of data collection, are not technologically savvy.

4. **Lack of Interaction among diverse data sets:** Creating effective new data systems within the NDEAR Ecosystem and integrating legacy data systems into the NDEAR ecosystem will enable the interaction of diverse data sets such as UDISE, Pratham's ASER, National Achievement Survey ('NAS')[39], and other data sets managed by various state governments. It is important to highlight that the NAS Survey released by the Department of School Education and Literacy, Ministry of Education is a school based survey, where as ASER is a household based survey, as it aims to include all children who may be enrolled in government schools, private schools or those who are not enrolled in school. At present there exists a level of distrust[40] in the findings of both surveys- ASER as well as NAS; as these surveys provide different results for similar metrics[41]. Though the findings of these surveys are critical to assess the learning outcomes of school students in India, there is no engagement between the organisations behind the survey—namely Pratham and the Ministry of Education. This highlights a missing opportunity of collaboration between the government and the private sector.

5. **Streamlining Data Collection and Management:** The current data collection model is labour-intensive and fraught with challenges, from manual data entry to the logistical nightmare of paper-based records management. This inefficiency is evident in the lengthy processes required for data submission, such as the UDISE+ reports, which demand extensive manpower and time, contributing to operational delays, data inaccuracies, and a compromised experience for all stakeholders. A proposed solution is the creation of a unified metadata structure that simplifies data handling at all

[39] National Achievement Survey "The National Achievement Survey (NAS) is a national level large-scale assessment conducted to obtain information about the learning achievement of students of Classes 3,5,8 and 10 studying in State Government Schools, Govt. Aided Schools, Private Unaided and Central Govt. schools. It is a national representative survey that provides a system level reflection on the effectiveness of school education."

[40] Disha Nawani, The Problem with ASER Survey, January 2023

[41] Muthyanolla, Sai Krishna. "Review: As NAS-2021 shows a decline in learning outcomes, here is what these surveys are all about." Factly, May 2022.

administrative levels, enhancing accessibility, and reducing the potential for errors and inefficiencies in the educational data ecosystem.

6. **Inadequate Incentive and Support for Teachers Assigned to Data Entry Tasks:** The task of making data entries at the moment is an additional administrative burden for teachers without any additional monetary compensation. This further drives them away from a meaningful engagement with data. While certain ICT workshops are held for capacity building, to enable teachers to make the data entry process but these do not provide them with any support during the process of data entry. This often results in errors and a sense of frustration among teachers.

7. **Challenge of Decentralisation and Fragmentation:** While the NEP has laid out a progressive framework, a significant challenge in its implementation lies in the decentralised nature of governance of education. Education, being in the concurrent list of the Indian Constitution, allows both the states and the central government legislative power over the sector. In the face of political shifts, this structure contributes to the challenges in implementing the NEP, as evidenced by the Karnataka state government's reversal of its commitment to the policy[42].

8. **Data Privacy and Security:** With the aggregation of vast amounts of student data, ensuring stringent data privacy and protection is paramount. With the involvement of multiple stakeholders, each with different levels of data access and handling capabilities, this becomes increasingly challenging. For instance, The Automated Permanent Academic Account Registry (APAAR) ID, proposed under the NDEAR framework, aims to securely store and manage student data throughout their academic life, tracking dropouts, academic performance, and extracurricular activities. However, it has faced criticism[43] for opacity, insufficient legal protections, and over-collection of data. A proposal to link these IDs to Aadhaar numbers without adequate public disclosure or clear data usage policies has raised privacy and security concerns, especially for children's data, and is said to be in conflict with Supreme Court directives

[42] Karnataka Government Scraps NEP

[43] Verma, Disha. Waghre, Prateek. "Aadhaar-based student ID raises alarms for privacy." Internet Freedom Foundation, October 2023.

on Aadhaar. The Internet Freedom Foundation has sought more information through RTI Applications, calling for legal safeguards and a transparent implementation of APAAR. An additional challenge has been the data security for individual data portals.

“Inaccuracy is alleged sometimes by extrapolating from non representative surveys or events in specific states, which is often demoralising for teachers. And then they are treated with doubt”

- Economics of Education Researcher

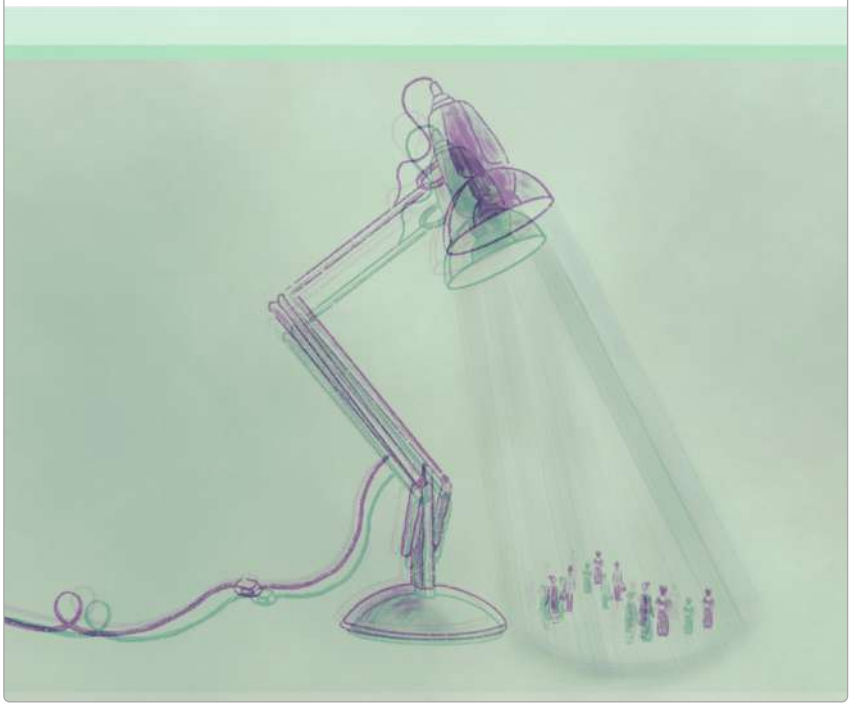
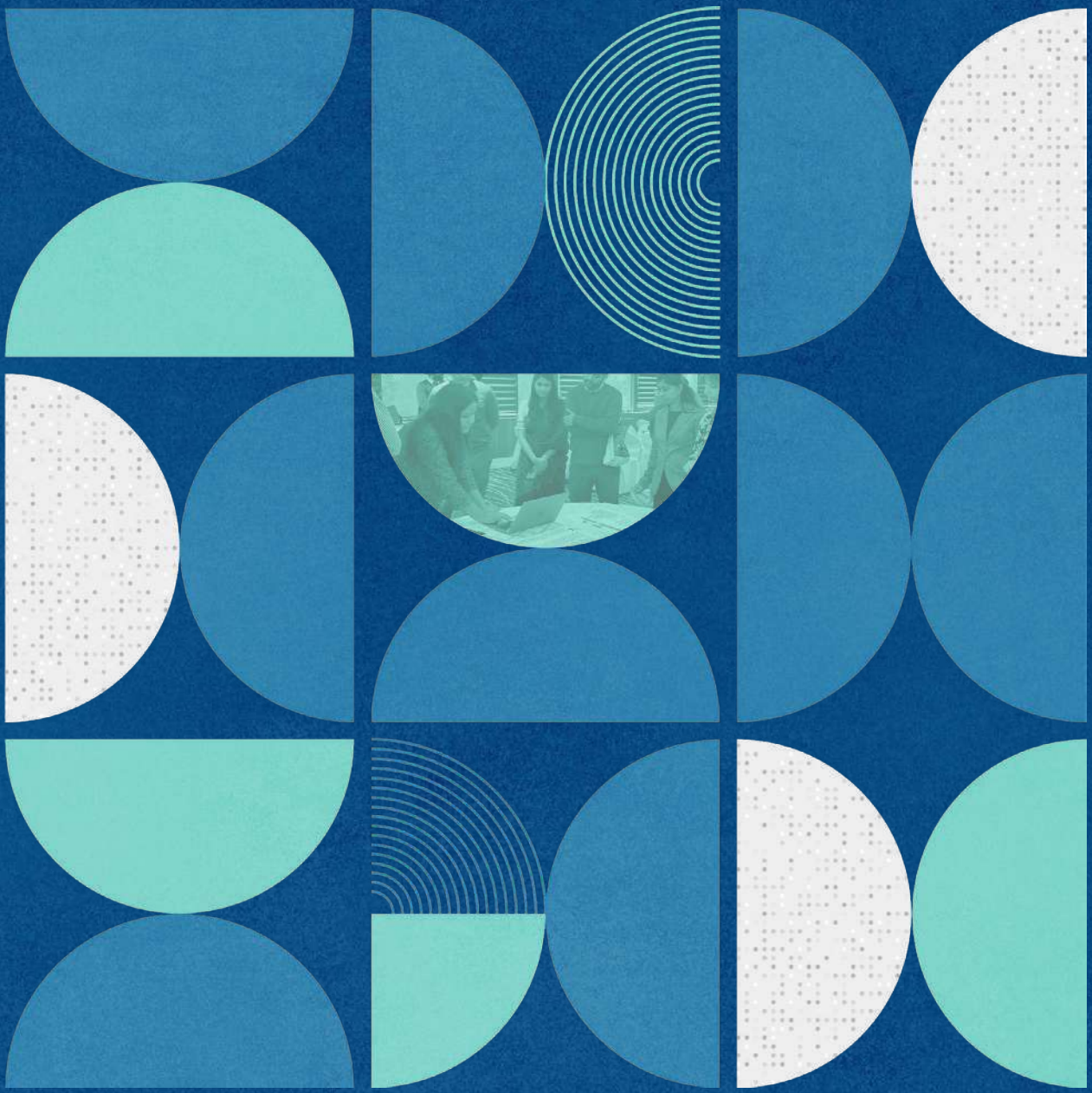


Image courtesy of Quicksand Design Studio

SECTION 6

Conclusion & Recommendations



SECTION 6

Conclusion & Recommendations

NDEAR's implementation, with its focus on interoperability and inclusivity, is pivotal for leveraging data in Indian schools for actionable insights. However, substantial challenges such as inadequate digital infrastructure, the transition to a dynamic, student-focused system, and data privacy persist. The working group's efforts to bridge the gaps in data collection and management, and to enhance data-mindedness and literacy among teachers and students, are essential for achieving universal access to quality education. These efforts also align with global standards and aim to ensure that no student is left behind in the rapidly evolving digital age. Through their project outputs—the publicly accessible report that promotes data as a tool for empowerment; a bilingual gamified learning platform (in English and Malayalam) designed to foster responsible data practices; visual narratives that depict the daily lives of teachers and data collectors; and a contextualised framework for assessing education data quality—they have highlighted critical issues in data management and usage in India. These need to be addressed before the Indian school education system can fully leverage technology for transformation. We at Aapti, believe that adopting the following recommendations would contribute significantly to strengthening the data ecosystems in education, ensuring data accuracy, enhancing stakeholder engagement, and ultimately improving educational outcomes.

1. **Establishing a Centralised Data Collection Agency with Open Data Principles:** We recommend the formation of a central agency responsible for the annual collection of educational data through a standardised and rigorous process, coupled with the promotion of open data principles. This dual strategy aims to avoid duplication of data requests and foster a sense of ownership and rigour within the data

collection framework, ensuring that once specific data is collected, it is not solicited again. By facilitating data sharing and integration across different government levels and agencies, this approach streamlines the process for stakeholders, reduces the administrative burden on teachers and educational administrators, and enhances both the efficiency and reliability of the collected data, promoting a cohesive and transparent educational data ecosystem.

2. **Providing Incentives for Teachers in Data Entry Tasks:** To enhance the effectiveness of data entry, it is essential to incorporate specific measures to support and incentivise the efforts of teachers. Teachers undertaking such tasks should receive financial incentives acknowledging the additional workload. To facilitate efficient and accurate data management, every teacher involved in data entry should be equipped with a dedicated workstation, including a laptop or computer with reliable internet access. They should have access to support staff via phone should they need assistance with challenges during the data entry process, whether they involve correcting entries or addressing system malfunctions. The adoption of dictation technology, where possible, is also recommended to streamline the data entry process, making it less cumbersome and more time-efficient.

3. **Deploying Data Specialists for Rural Data Collection:** To bridge the technological divide and enhance data collection accuracy and efficiency, particularly in rural areas, we recommend having data specialists in each administrative block. These specialists will provide essential support to schools by facilitating the use of technology in data collection processes. To realise this objective, a comprehensive training program should be developed, focusing on the nuances of educational data management, effective application of technology in educational settings, and strategies for engaging with local communities. Additionally, it is crucial to secure funding to support the recruitment and deployment of these data specialists across various blocks. This initiative addresses the critical challenge faced by rural schools, which often lack personnel proficient in technology, thereby ensuring more precise data collection and fostering greater accountability in relation to data collection.

4. **Strengthening Infrastructure and Capacity:** The present digital infrastructure is inadequate[44] and provides poor support for efficient data handling, as evidenced by subpar internet services and a lack of proper IT training for school staff. We recommend strengthening the existing infrastructure, reallocating budgets more effectively towards annual training of IT staff and teachers, and ensuring reliable internet connections. Essential measures include provisioning fully operational computer systems, prioritising the deployment of a 5G network in schools nationwide and utilising the Gati-Shakti along with GIS mapping, which will lay the groundwork for a comprehensive, robust infrastructure. This recommendation addresses the critical issues identified during the SDE workshop, such as poor internet connectivity, and sets a strong foundation for future technological advancements in Indian schools.

5. **Culture of Feedback and Iterative Development:** To achieve policy goals with the emerging NDEAR architecture, we advise establishing a robust, bottom- up feedback loop. This mechanism should comprehensively assess the relevance and utilisation of collected data, alongside operational metrics such as data entry frequency and regional update discrepancies. By pinpointing underperforming areas, this approach will enable targeted interventions and the formulation of bespoke solutions. Crucially, the feedback system must involve teachers directly to understand their concerns, channelling these insights to block education officers or district officials for responsive action. This adaptability will be crucial, accounting for regional variances and incorporating insights from a broad spectrum of users. Such an approach ensures that the feedback loop not only identifies the gaps but also encourages a culture of continuous improvement and responsiveness.

6. **Addressing Funding Issues in Data Governance:** While recommendations to enhance the digital infrastructure are critical, it is equally important to give agency to the stakeholders, such as teachers giving them the autonomy to make changes. To achieve this, various NGOs and private organisations could consider a pivot in their funding strategies, focusing on the impact of their data (i.e. ability to influence systemic change over the long term) and not just the outputs (i.e. immediate results). For example, Brazil's "one laptop per child"[45]

[44] As per the data available on UDISE+ dashboard on infrastructure and facilities, out of 1.49 million schools only 504,989 have access to internet; 707,498 have access to computers and only 1,289,245 have functional electricity.

[45] Adi Robertson, 'OLPC's $100 laptop was going to change the world — then it all went wrong', The Verge, April 2018.

highlighted the need for educational technology investments to include measures that affect learning outcomes, such as reading and writing skills, rather than merely providing devices and teaching children how to use them. Data Governance establishes a strategy for enacting the role that data can, should, and is obligated to play, in achieving results.

7. **Incentivising Compliance to Maintain Data Standards:** Linking financial incentives, particularly through schemes like POSHAN, can motivate state governments to align with central guidelines and the NEP. If funding can be tied to the quality of data management, this could further incentivise the state governments to invest in the necessary infrastructure and processes.

8. **Creating Leaderboards for Benchmarking Schools:** We recommend developing a standardised scoring system for benchmarking schools. This would allow schools to recognise their effectiveness in utilising the collected data.

> **"They will immediately ask for government data to be entered in 2-3 days. It is not difficult work, we should just be given sufficient time to do it. Teachers have to teach. Non-teaching staff have other work too – they cater to the parents who come to pay fees, for other forms, bonafide certificates etc. Data work has to be adjusted in the limited time we have."**
>
> **- IT Staff, Private School**

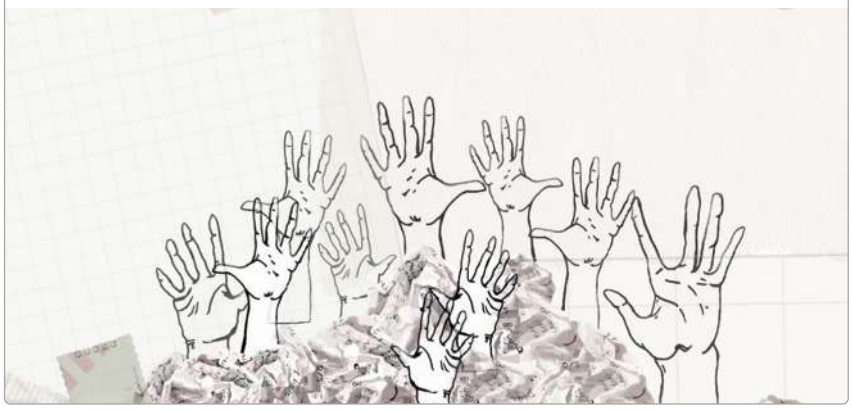
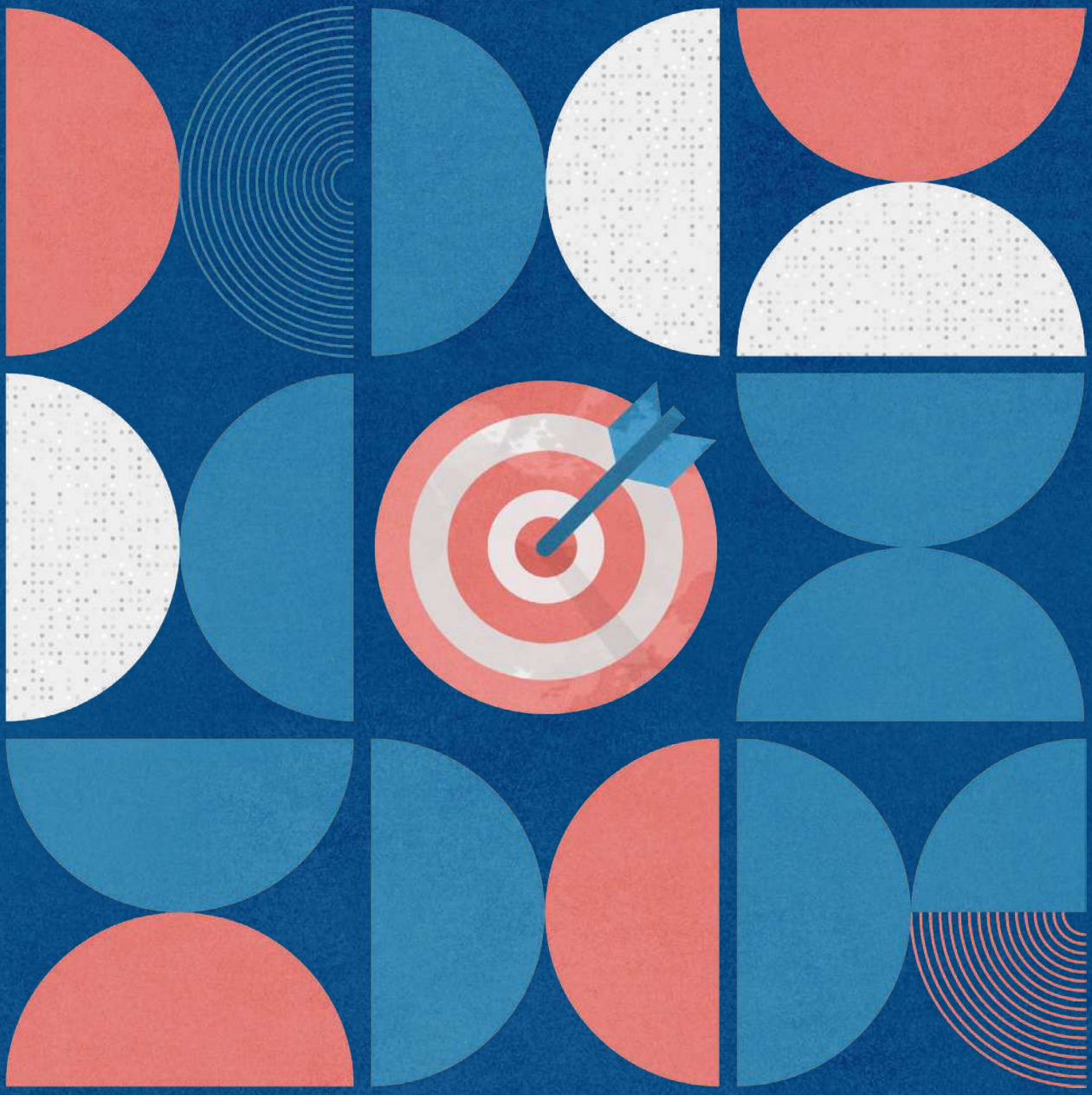


Image courtesy of Quicksand Design Studio

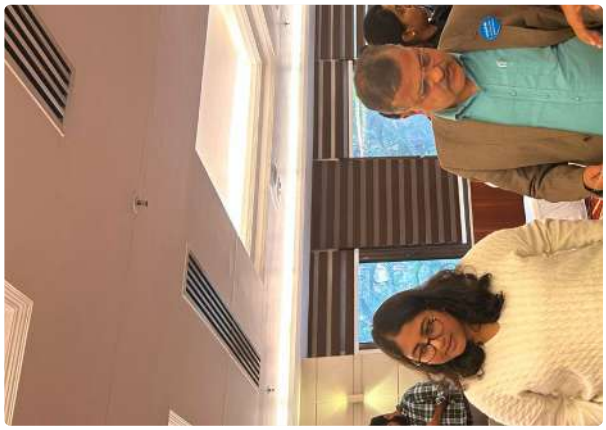
SECTION 7

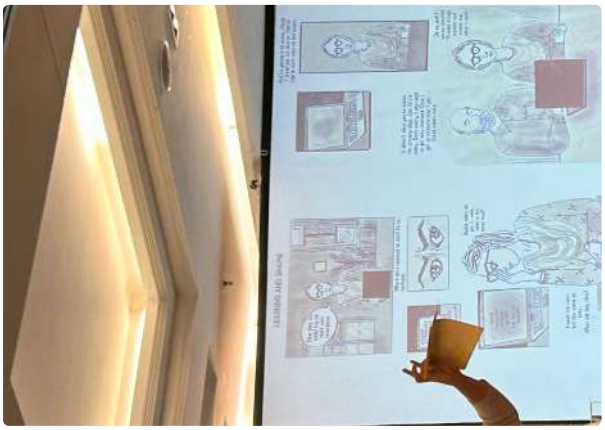
Glimpses from the Event

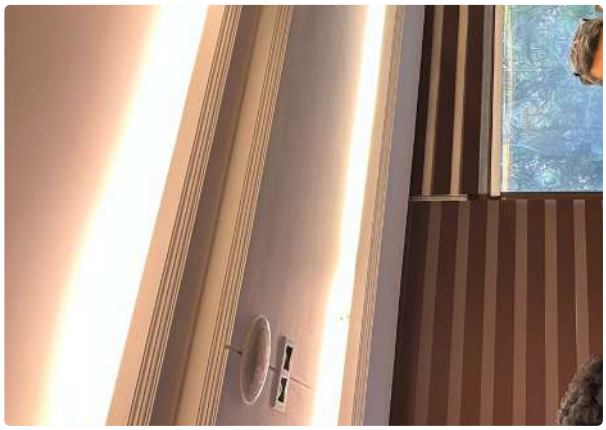


SECTION 7

Glimpses from the Event







SECTION 8

Annexures

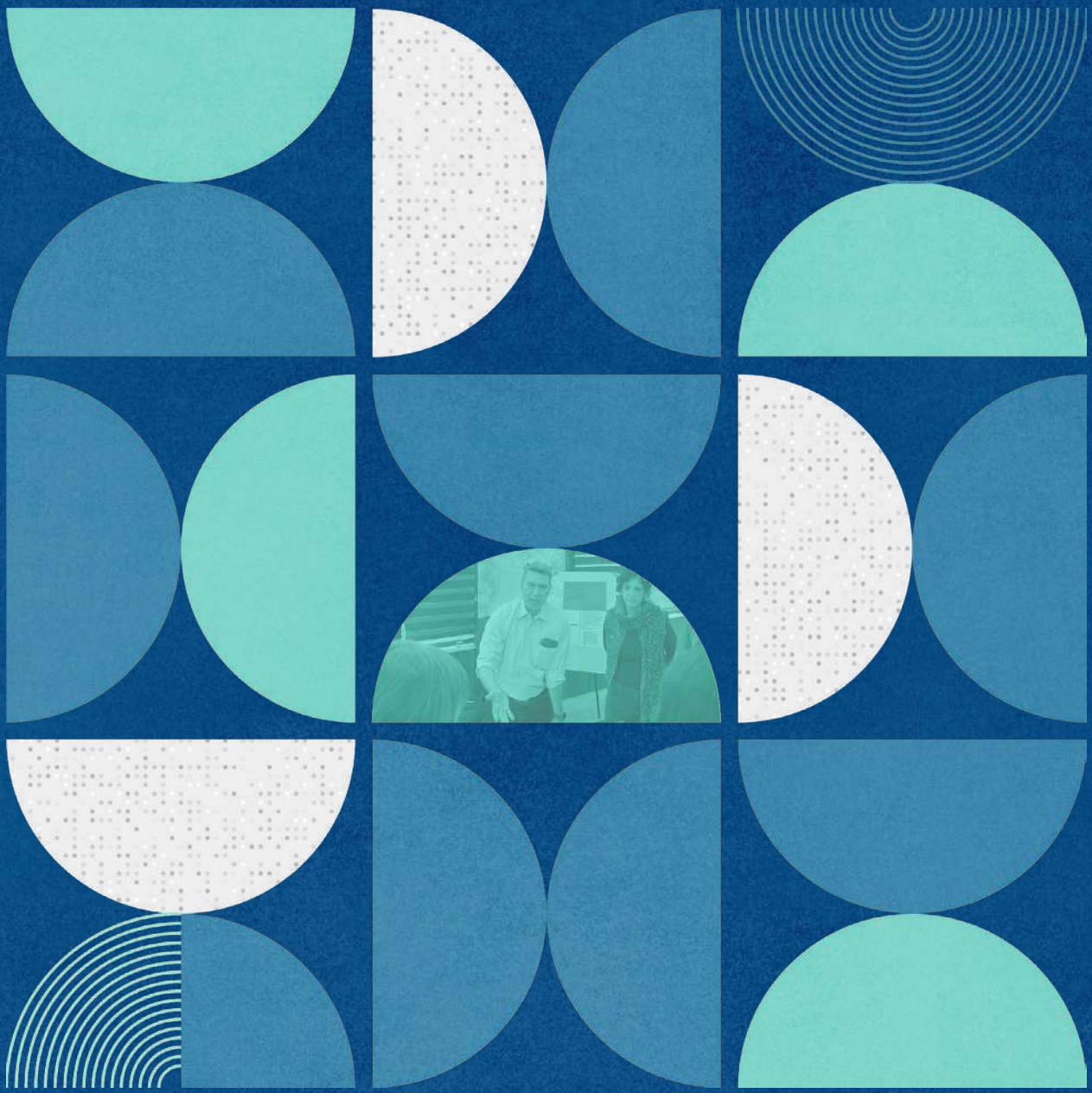


TABLE OF CONTENTS

SECTION 8

Annexures

Detailed Agenda of the SDE Workshop

ANNEXURE 1

Workshop on Strengthening Data Ecosystems (SDE) in Indian Schools

Dates: 27-30 November 2023
Venue: India Habitat Centre. Lodhi Road, New Delhi - 110003
Tamarind Conference Room

Date & Day	Event Title	Event Description	Start time	End time
*Day 1				
Monday, 27 November (closed group)	Registration and Networking		9:30	10:00
	Welcoming the participants	A short session led by the Aapti team welcoming the participants to the workshop. Quicksand to run through the agenda briefly	10:00	10:15
	Ice Breaker & Tea Break		10:15	10:45
	Keynote: Shaping the future of Data Ecosystems in Education: Insights from USAID	Representatives from USAID share their perspectives on the importance of strengthening data ecosystems in Indian schools. Genesis of the SDE initiative- the goals and the long-term vision. Includes a Q&A, allowing attendees to engage directly with the speakers	10:45	11:30
	Group Work: Key challenges in School	Mixed groups of participants to share biggest issues in data	11:30	12:30

	Data Ecosystems	ecosystem in education and how seek to address it		
	Share Back		12:30	12:45
	Lunch Break		12:45	13:30
	Working Group Presentation Data for Education Systems in India: Design, Processes, and Use Q&A	A WG presentation, the attendees will be given a worksheet to write down reflections and ideas	13:30	14:15
	Working Group Presentation Stor(y)ing Data Inaccuracy: Detechnicising our imagination of data production Q&A	A WG presentation, the attendees will be given a worksheet to write down reflections and ideas	14:15	15:00
	Tea Break		15:00	15:30
	Working Group Presentation Ed-DataQuality+: Empowering Education Decision-Makers with data quality assessment Framework Q&A	A WG presentation, the attendees will be given a worksheet to write down reflections and ideas	15:30	16:15
	End of Day Reflections	Quick reflections on the day. Give the participants a small task to share back on Day 2	16:15	16:45

*Day 2				
Tuesday, 28 November (closed group)	Data Stories	Quick share out on a wall of participants' data stories	10:00	10:30
	Shaping the future of Data Ecosystems in Education: Insights from Mozilla Foundation	Mozilla Foundation speaker(s) to share insights around data sovereignty and how school ecosystem data aligns with these principles	10:30	11:15
	Working Group Presentation	A WG presentation, the attendees will be given a worksheet to write	11:15	12:00

	Towards Enhancing Data Utility and Accessibility for Schools Q&A	down reflections and ideas		
	Lunch Break		12:00	13:15
	Working Group Presentation Enabling the development of localised understanding of data as a tool for empowerment in Indian schools through collaborative and story-based pedagogy Q&A	A WG presentation, the attendees will be given a worksheet to write down reflections and ideas	13:15	14:00
	Shared Reflections	A facilitated exercise to discuss and reflect on insights from the 5 working group presentations	14:00	14:45
	Tea Break		14:45	15:15
	Design Exercises	Engaging participants in creative thinking to identify and understand the needs of end users of the working group's outputs	15:15	16:15
	Preparation for the upcoming exhibition		16:15	17:15

*Day 3				
Wednesday, 29 November	Fireside chat: **Enhancing Learning Outcomes through Implementation of Educational Technology in India**	With external speakers	10:00	11:30
	Tea Break		11:30	12:00
	Paired Reflections	Paired discussions to share key learnings and reflections, drawing from the presentations and mutual experiences	12:00	13:00
	Lunch Break		13:00	14:00
	Unfacilitated working	Allocated co-working period for	14:00	18:00

	time for the WGs and set up for Day 4	working group members to collaborate, deepen their understanding, and share research, along with preparation for Day 4		

*Day 4				
Thursday, 30 November (open to invited public)	Opening panel discussion (public) -**Unlocking the Value of Education Data for Societal Good: Shifting from Data Collection to Value-Oriented Governance**	With external Speakers	10:30	11:15
	Walk through of projects, networking and coffee (public)	Exhibition: Print & screen based	11:15	12:00
	Panel Discussion (public)	A panel discussion featuring a representative from each WG, moderated by Aapti	12:00	13:00
Thursday, 30 November '23 (closed group)	Lunch Break		13:30	14:30
	Closing Reflections & Taking the work forward	A concluding session to reflect on key takeaways and strategize future directions for the initiative	14:30	16:00

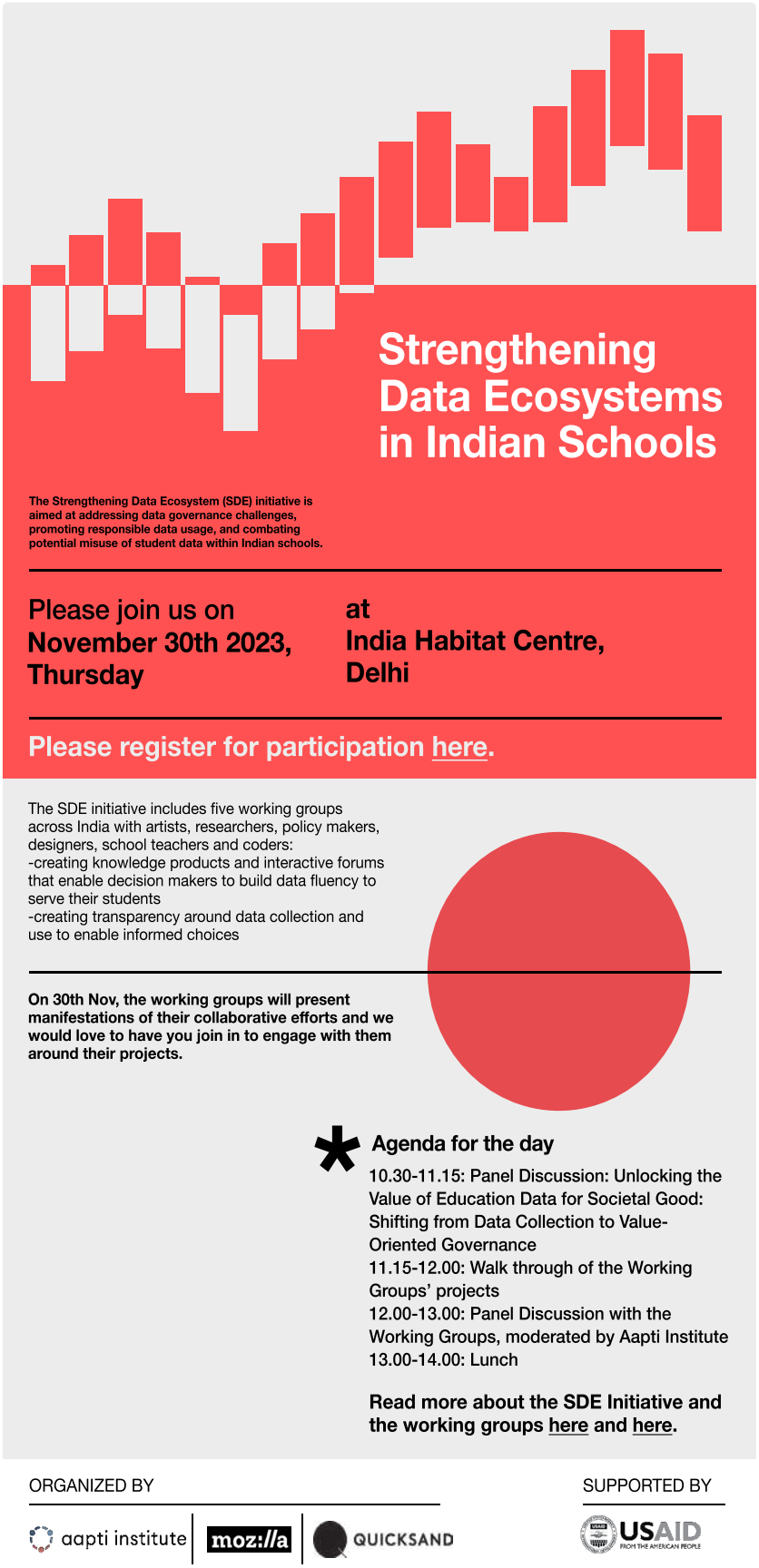
ORGANIZED BY

aapti institute | moz://a | QUICKSAND

SUPPORTED BY

USAID FROM THE AMERICAN PEOPLE

ANNEXURE 2


Strengthening Data Ecosystems in Indian Schools
The Strengthening Data Ecosystem (SDE) initiative is aimed at addressing data governance challenges, promoting responsible data usage, and combating potential misuse of student data within Indian schools.
Please join us on November 30th 2023, Thursday
at India Habitat Centre, Delhi
Please register for participation here.
The SDE initiative includes five working groups across India with artists, researchers, policy makers, designers, school teachers and coders:
-creating knowledge products and interactive forums that enable decision makers to build data fluency to serve their students
-creating transparency around data collection and use to enable informed choices
On 30th Nov, the working groups will present manifestations of their collaborative efforts and we would love to have you join in to engage with them around their projects.
Agenda for the day
10.30-11.15: Panel Discussion: Unlocking the Value of Education Data for Societal Good: Shifting from Data Collection to Value-Oriented Governance
11.15-12.00: Walk through of the Working Groups' projects
12.00-13.00: Panel Discussion with the Working Groups, moderated by Aapti Institute
13.00-14.00: Lunch
Read more about the SDE Initiative and the working groups here and here.
ORGANIZED BY
aapti institute
mozilla
QUICKSAND
SUPPORTED BY
USAID

ANNEXURE 3


Dates: 27th-30th November
Strengthening Data Ecosystems in Indian Schools
The Strengthening Data Ecosystem (SDE) initiative is aimed at addressing data governance challenges, promoting responsible data usage, and combating potential misuse of student data within Indian schools.
ORGANIZED BY
aapti institute
moz://a
QUICKSAND
SUPPORTED BY
USAID
FROM THE AMERICAN PEOPLE

aapti institute

Aapti is a public research institute that works at the intersection of technology and society. Aapti examines the ways in which people interact and negotiate with technology both offline and online.

contact@aapti.in | www.aapti.in