OEA Digital Access

Use Case Defined

March 2022

1) The Use Case Problem

Defining the Problem: What problem does this use case seek to solve?

- **Digital Access**: How does an education system ensure all students in their system have access to devices and digital learning applications and platforms when they are learning remotely, for example from home or outside of a physical school?

During the 2020-2022 Global COVID-19 Pandemic, as schools shifted to remote learning, it became clear that large numbers of students did not have access to digital learning from their homes or outside of schools. Closing "Digital Equity Gaps" became the immediate focus of many school systems around the world, as these gaps were largely defined by socio-economic, racial, immigrant status, and gender factors. The pandemic magnified existing disparities. While investments and progress have been made in closing these digital access gaps, there are still many student populations who do not have technology resources to effectively learn using digital tools. Those who do not have access, especially outside of physical schools, continue to be disadvantaged in learning opportunities.

Education systems seeking to close digital equity gaps need to use limited resources very efficiently to provide digital learning access to all students. Microsoft Education, Kwantum Analytics, and Fresno Unified School District in California collaborated to create an open source GitHub module on [Open Education Analytics (OEA)](#) to enable education systems to use data effectively see which students have digital access for learning and which ones don't, so that they can ensure every student in their schools has access to learning whether they are in school or learning from home or outside of a physical school.

2) The Use Case Stakeholders

Considering the benefits of a use case – and its potential harms - requires the consideration of different stakeholders and their points of view. Stakeholders typically include the people who are responsible for, will use, or will be affected by the use case. Stakeholders are defined by their role: their duties, contextual identity, or circumstances in relationship to the use case.

- **Direct stakeholders** interact with the data from a use case directly and make decisions or take actions based on results. They include data users, system developers, and even data system maintenance staff.
- **Indirect stakeholders** are affected by the use case data system but, unlike direct stakeholders, they do not have a role that requires them to use or maintain the system. Indirect stakeholders can include groups who may be affected by the downstream effects of the system, such as parents, students, or future employers.
- **Malicious actors** include hackers and others who may intentionally misuse the system. Considering malicious actors is important to supporting safe and reliable data systems.

Who are the stakeholder groups for this use case, and how are they involved in its development?

Possible Stakeholder Groups	Relationship to Use Case	Involvement in Use Case
Students	Indirect Stakeholders	Potential beneficiaries of new devices, better connectivity, or IT training or support if they are identified as having no or low access.
Parents or Guardians	Indirect Stakeholders	Potential beneficiaries of new devices, better connectivity, or IT training or support if their students are identified as having no or low access.
Educators (Faculty or Teachers)	Indirect Stakeholders	May use data indirectly to understand whether all students in a class or course can learn using digital tools outside of school (e.g., to assign homework that requires connectivity).
Technology departments and staff	Direct Stakeholders	Will use data directly to provision devices, connectivity, IT support, and training to students and families with no or low digital access outside of school
School System or Institutional Leaders	Indirect Stakeholders	May be needed to provide funding for devices or to lobby Internet service providers
Funding Groups and Internet Service Providers	Indirect Stakeholders	May be involved to receive data outputs from the use case, to

		provide funding or connectivity solutions
Potential Malicious Actors	Indirect Stakeholders	Potential for hacking this data system to identify locations for stealing devices outside of schools

Engaging stakeholders in the use case development process.

Including stakeholders in the early thinking and conceptualization of a data use case is a good way to ensure that the use case output will be accepted, trusted, and used by key stakeholders. For example, conducting interviews or focus groups with representatives of each stakeholder group can provide early insights into the conceptual model framing the use case and the appropriateness of specific data sources to be used. At a later stage in the project, involving key stakeholder groups in designing the use case outputs (such as a dashboard or notification) can be essential to that product's eventual effective use.

Outline how stakeholders will be involved in the development in different stages of the use case development:

At an early stage in the use case process the primary stakeholders involved should be the technology team and school/system leaders. The project will likely be led by the technology team, with leaders included to define the project scope to ensure those leaders can and will act on the data to provide technical or budget resources to address digital equity gaps. In some cases, local or national funding groups and Internet service providers may also be included as stakeholders.

At a later stage in the use case project, as dashboards and information start to be available, these should be reviewed for validity with different stakeholder groups. For example, students (and their parents/guardians) who are identified as having no or low digital access should be contacted to check whether the data is correct, or if students actually do have access to digital learning tools that have not been picked up by the data sources used in the project.

Leaders, funding groups, and Internet service providers should also be consulted to ensure that the data outputs (e.g. dashboards) are intuitively understood, seen as legitimate, and can be used for decision-making and tracking progress.

What type of outputs are expected from this use case, such as AI models, dashboards, or notification systems?

This use case should result in a data dashboard that shows patterns of digital access and identifies which students have no or low connected devices outside of schools. This can allow education system leaders to target their resources (e.g. provision devices, provide internet connections) most efficiently to ensure all learners have digital access to learning outside of school.

3) Mapping Theory to Data

For this use case, what prior research or conceptual model frames your theory of the problem?

Access to digital learning resources outside of school requires a number of key factors:
1) Devices
2) Connectivity: access to the Internet or cellular services
3) Digital literacy to be able to use digital learning tools effectively
4) Technical support to ensure digital tools function

Digital learning resources that can be used on devices are also necessary to effective digital learning. Engagement with digital learning resources will be addressed in a separate Open Education Analytics use case on digital learning engagement. The use case for this document describes only access to digital learning (devices and connectivity).

- Drane, C.F., Vernon, L. & O'Shea, S. Vulnerable learners in the age of COVID-19: A scoping review. Aust. Educ. Res. 48, 585–604 (2021). https://doi.org/10.1007/s13384-020-00409-5
    - "A lack of digital skills and digital access can have a negative impact on learning."
    - "As evident in the Australian Digital Inclusion Index (Thomas et al. 2019) which measures digital inclusion in three discrete ways (access, affordability, and digital ability), a digital divide exists between students from low and high socio-educational backgrounds."

- Chandra, S., Chang, A., Day, L., Liu, J., McBride, L., Mudalige, T., Weiss, D., Fazlullah, A. Closing the K–12 Digital Divide in The Age of Distance Learning. (2020) https://www.bbcmag.com/pub/doc/BBC_Nov20_DigDivide.pdf

    - "Our analysis finds that for a robust distance-learning experience, students and teachers need four things: high-speed internet service, internet-enabled devices that support assignment completion (excluding cell phones), distance learning instructional content, and support, including digital literacy, teacher readiness and technical support."

    - "Closing the digital divide will require public and private sectors to come together with a sense of urgency for immediate action to ensure equitable learning opportunities during the pandemic and a sustained commitment to secure the nation's educational future by ensuring that digital technology will benefit all students and their families."

- Workie, Essey; Hinkle, Lillie; deDufour, Anna; Lacarte, Valerie. Advancing Digital Equity among Immigrant-Origin Youth. Migration Policy Institute. https://www.migrationpolicy.org/sites/default/files/publications/mpi-digital-equity-2021_final.pdf
  - "Digital tools are central to performing daily tasks--from completing school assignments and applying for jobs, to reviewing personal health records, and making financial transactions."
  - "These obstacles [lack of digital access and training] were reported to have led to knowledge gaps, lower grades, chronic absenteeism, and disenrollment."

Mapping Theory to Data. From prior research or conceptual models what are they key data categories expected to inform this use case? What local data sources are available or needed for each category? Please note where no data is available for a Data Category

A key part of the use case development process is deciding which data to use and how it should be mapped to the theory of the problem. Identifying which data should be viewed as a "feature" and which data is the "target outcome" is at the core of this mapping.

Key Data Category	Local Data Source
1. Device Access	If devices are provided by an education system to students, the device assignment data could be used. If devices are not provided directly, but students use their own devices that authenticate or login to the education system's digital platforms, then that identity data can be used.
2. Connectivity	For example, data from local internet provider on Upload/Download speed, latency, or request processing time, and general location data (in school or out of school location, for example). Also, timestamp data of connection and digital activity is needed (e.g., during or after regular school hours/day).
3. Student Information System or Management Information System	School System's SIS to provide total population of students who should be in analysis, along with key factors (e.g., grade levels, school affiliations).
4. Digital Literacy	Any data on training of students on how to use digital learning tools (not available with Fresno use case).
5. Technical Support Access	Data from a school system's technical help desk for students and families.

Note: OEA Modules and Packages, support many of these education use cases through accelerating the ingestion of key data sources needed and providing resources to set up these use cases.

OEA Modules that can be used for this use case:

OEA Module	Description
Ed-Fi Data Standards	For typical Student Information System (SIS) data, including school rosters, grade level and demographic information.
Connectivity Data	Microsoft Graph (SignInAuditLogs query) data.
Digital Learning Apps and Platforms	Microsoft Education Insight Premium, Clever for learning application data and iReady for language and math assessments and learning activities.

Note: Mapping theory to data with a 'data dictionary.'

A "data dictionary" allows the data team to examine specific data tables and data entities in the available datasets, and then map specific items to the Key Data Category.

New data services like Azure Purview can support this work through creating a holistic, up-to-date map of a data repository with automated data discovery, sensitive data classification, and end-to-end data lineage.

Please see "Privacy and Security" section below for more ensuring that sensitive data is protected.

4) Responsible AI Principles Applied

In these next sections, please answer the questions under each of the headings describing how responsible AI principles will be applied to this use case.

Fairness Principle

AI systems should treat everyone in a fair and balanced manner and not affect similarly situated groups of people in different ways. Human decision makers are susceptible to many forms of prejudice and bias, such as those rooted in gender and racial stereotypes. To ensure AI models are trained in a way that does not embed or re-enforce those biases, models must be tested for fairness. Microsoft has developed an open-source toolkit to support this called Fairlearn, which can be applied within the Azure analytical services used in the OEA reference architecture.

Video on Fairness Principle.

Who is most likely to be at risk of experiencing harms from this use case?

This Digital Access Use case does not involve the development of machine learning or a predictive model, so AI model bias is not applicable. However, data that describes different groups of students or contexts should be used to understand patterns of bias in the distribution of digital access resources.

Group or Subpopulation	Clearly Labelled in Dataset? Y/N	Planned Mitigations
1. For example, immigrants or rural students		Provide dashboard charts and data that show if students in this group have higher or lower levels of digital access
2. For example, socio-economically disadvantaged, gender, language-based groups, racial, ethnic, or religious groups		Provide dashboard charts and data that show if students in this group have higher or lower levels of digital access

Reliability and Safety Principle

Systems should operate reliably and safely when they function in the world. AI systems must be designed with a view to the potential benefits and risks to different stakeholders and undergo rigorous testing to ensure they respond safely to unanticipated situations and do not evolve in ways that are inconsistent with the original shared purpose.

Video on Reliability and Safety Principle.

What are possible risks faced by learners or educators from the analytics of this use case?

1) Risk 1: students who do not have digital access outside of school who are not identified in the system.

2) Risk 2: students who do have digital access outside of school but are identified as not having access.

3) Risk 3: teachers who believe their students have digital access outside of school and assign learning that relies on that access, when there are some students who still do not have access.

Planned Mitigations:

The data solution for digital access needs to be made as robust and accurate as possible, and repeated, continuous human checks of the system conducted. For example, schools or teachers can confirm directly with parents or guardians whether their students have digital access when they are attending school in person. Technical teams can work to ensure school and class rosters are accurate, and that no student is missed in the analysis.

Transparency Principle

Transparency requires visibility into all levels of decision-making and design of an AI system. Designers should clearly document their goals, definitions, and design choices, and any assumptions they have made. Those who build and use AI systems should be forthcoming about when, why, and how they choose to build and deploy them, as well as their data and systems' limitations. Information should be readily available on the quality of the predictions and recommendations the AI system makes. Transparency also encompasses intelligibility, which means that people (in this case, educators, parents, students, etc.) should be able to understand, monitor, and respond to the technical behavior or recommendations of AI systems.

Video on Transparency Principle

What steps will the analytics or AI process include?

This Digital Access Use case does not involve the development of machine learning or a predictive model, so AI transparency on model limitations is not applicable.

Who will develop the analytics or models?

Not applicable for Digital Access as no machine learning modelling is conducted. For the Fresno Unified Digital Access dashboard examples, Kwantum Analytics developed the data exploration dashboards.

How will the limitations of the analytics or AI model be communicated to stakeholders and users?

Limitations on the quality of the data sources should be noted as information within the Digital Access dashboards themselves. For example, if a complete inventory of student devices is not available.

What means will be built into the system for correction and model feedback by those who provide data and who use its outputs?

An email address and system-level point of contact for providing feedback to the system should be posted within the Digital Access dashboard directly.

Privacy and Security

Private or personal data should not be collected or incorporated in analytics or AI products for education unless all groups have agreed this data is necessary to achieve the shared purpose of a specific analytics or AI project. Additionally, the people providing the data need to give permission for the data to be used for this purpose, such as through school policy at enrollment. Ideally, data providers should directly understand the value that they will receive as a result of sharing their data. Finally, the security of that data must be protected, guidelines or policies developed for which roles can access which data, and the level of anonymization needed for specific use case purposes defined.

[Video](#) on Privacy and Security principle

Identifying sensitive data, such as personal information, should be part of the Digital Access use case process. For individual datasets, sensitive data should be pre-identified, and scripts written to pseudonymize or anonymize specific data fields before they "land" in Stage 2 data lakes and are accessed by researchers or data scientists. Only this protected data is pulled into Stage 3 data lakes for presenting in dashboards.

How will access to sensitive data be secured and protected in the data environment?

The Digital Access dashboards will only be available to those stakeholders who should have role-based access to such information, implemented through Azure Active Directory, and as governed by the system's policy for data governance.

Open Education Analytics: Data and AI Use Case Template

Does the dataset contain any personally identifiable information (PII) and how will that data be protected and governed?

The Digital Access package includes datasets that contain PII in the Student Information System. That data will be protected by pseudonymizing this dataset in all dashboards and published data.

Accountability

Accountability requires that people who develop and deploy AI systems be held responsible for how they operate. AI systems should never be left to operate unchecked, irrespective of the degree to which they may be capable of acting autonomously. This is what is meant by the phrase "humans in the loop." A part of this is ensuring documentation of the decisions made during the AI system development. This document can be used for that purpose.

Video on Accountability principle

Who is responsible for reviewing the Use Case documentation and ensuring that the implementation meets responsible AI principles?

The decision makers (primarily the technology department) in the education system who use the Digital Access dashboards in practice to identify supports for students and schools will be responsible for continued implementation of the principles responsible data analytics described in this document. They should review this documentation thoroughly and update it if decisions or data changes.

How will stakeholders and end users be trained on the appropriate use of the system?

This document should be maintained as a reference. Education system leaders and schools will be responsible for training schools, educators, and all support staff and stakeholders on how to understand the dashboards, and on their intended use to inform decisions and actions on Digital Access.

How will the analytics or AI system be monitored over time to ensure analytics and prediction perform reliably? Who will be responsible for this?

Not applicable for this Digital Access use case as no machine learning modelling is conducted.

Inclusion

The datasets used in learning analytics and AI determine the insights and predictions produced. If those datasets do not represent the whole population of learners, if the data quality is poor, or if certain types of data are not included in the models, it will decrease the accuracy, validity, and inclusiveness of the insights. Similarly, if the way the insights are acted upon by the system do not include all groups (e.g., students with disabilities), it can reinforce exclusion from learning opportunities.

Video on Inclusion principle

What are the constraints of these local data sources for this specific use case?

In the case of Digital Access dashboards developed with Fresno Unified, extensive data on all student sub-populations was available through the Student Information System dataset.

Dataset Name	Constraints or Limited Representativeness
1. IT Support and Training	For the Digital Access dashboards developed with Fresno Unified, IT support and training data was not available to be included in the dashboards. That data represents a factor identified through prior research to be associated with digital access.