Actual School Connectivity - Model Overview

The The Actual School Connectivity Model has a few high-level steps:

1. Load in local data - both geospatial and general
2. Load in technology and infrastructure data
3. Compute school and community bandwidth requirements
4. Select technology
5. Compute costs

Each of these steps has a few different paths that can be taken depending on the source data available, connectivity technologies available in a region, etc. When additional paths are needed, this analysis tool will continue to be updated.

1. Load in local data
    a. School Data: Provided by the education ministry or other government agency.
        i. R - required
        ii. O - optional
        iii. S - suggested for best accuracy

Need	Name	Input	Description	If Missing?
R	Lat	decimal	School locations - Latitude	Won't run
R	Lon	decimal	School locations - longitude	Won't run
O	School Name	string	School Name/ID	Random ID generated
S	Distance to Nearest Fiber	decimal	Distance from school to nearest fiber line - straight line distance.	Fiber won't an available option
S	Type of Cell Coverage	2G, 3G, 4G, 5G	Available cell coverage (2/3/4/5G). Blank if none.	Cell won't be an available option
S	elec_grid	Y/N	Is the school connected to the electric grid? The current version of the model does not incorporate the grid capacity or stability.	Assume all schools need solar

S	generator	Y/N	Does the school have a generator on-site? The current version of the model does not incorporate generator capacity and assumes that the generator has enough capacity to run any new and existing hardware.	Assume all schools need batteries
S	num_students	Integer	Number of students enrolled at school	Calculated based on local population
S	num_teachers	Integer	Number of teachers working at school	Calculated based on local population
S	num_classrooms	Integer	Number of classrooms being taught at any given time	Calculated based on number of teachers and students

b. Population Data: UN adjusted dataset. The GEOTIFF file is a 1 pixel - 100 meter resolution. Each pixel essentially has a reference latitude and longitude (center point), spans 50 meters in each direction, and returns the population within that pixel.
    i. Geotiff file for country available for download here: https://www.worldpop.org/geodata/listing?id=69
c. Project Specific Inputs: These inputs can be customized to the country, but have a set of representative default values that can be adjusted based on local feedback. They take the following inputs, in JSON format:
    i. Configuration:
    ii. Usage

EMIS_allowableTransferTime	Assuming that EMIS transfers can happen after school hours, the maximum amount of time that the transfer can take. Set based on the battery lifetime of a laptop, assuming that grid and backup power is lost immediately after starting the upload.
allowableWebsiteLoadingTime	Maximum time in seconds that a data-rich static website can take to load.

Allowable Document Loading Time	Maximum time in seconds that a complete data and image rich Google Document can take to load.
Allowable Completed Assignments Loading Time	Maximum time in seconds that a single student assignment can take to load.
Peak Hours	Number of hours of heavy usage by students and staff members.
Size of Website	Average size of a website (KB)
Size of Document	Average size of a Google Doc
Google Docs Bandwidth	Bandwidth required for responsive use of Google Docs
Internet Browsing Bandwidth	Bandwidth required for minimum responsive internet use
Video Data Rate (480p)	Bandwidth required to stream 480p video
Teacher Prep Hours	Hours per day that teachers spend prepping for class

iii. Assignments:

Student Prep Time	Number of hours that students spend outside of homework time doing research, reading materials, etc.
Size of Teacher Research Time	Number of hours that a teacher spends researching and preparing to create a single assignment.
Number of Daily Assignments Per Student	Number of assignments that a student must complete each day (averaged).
Student Research Time	Number of hours a student spends researching specifically to complete an assignment.
Student Assignments Time	Number of hours that a student spends writing up or completing an assignment.
Time to Grade One Assignment	Number of minutes that a teacher spends grading an assignment.

iv. Community:

Fraction of Community Using School Internet	Fraction of the community that uses the school-provided internet on premise ever. This is not a daily usage, but a TAM.
Session Length	Number of minutes that a community member is connected and using the school internet.
Weekly Sessions	Number of times that a community member comes to the school.
Community Access Hours	Number of hours that school bandwidth is opened up to the community for general browsing. May overlap.
Contention	Number of people sharing a single “slot” of bandwidth. Typical ranges are 3-40 depending on locale.

v. Lesson Planning

Weekly Planning Time	Number of hours per week a teacher spends making new lesson plans
Fraction of Planning Time Browsing	Fraction of lesson planning time that a teacher spends performing research and looking up new material. The remainder of the time is assumed to be spent researching offline, writing plans and creating materials, etc.

vi. EMIS - “Education Management Information System”. EMIS data is entered as a Pandas Dataframe Series, with the following inputs:

1. Mean Value
2. Minimum value (3 SD below mean) as a percentage discount from mean
3. Max value (3 SD above mean) as a percentage added to mean
4. Frequency with which this data is uploaded (times per year),
5. Unit of measure - the group of users across which the data is being measured

Admin: Enrollment	Student population and census data
Admin: Cohort	Class-by-class (cohort) data
Admin: Behavioral	Disciplinary records, etc
Admin: Special Needs	

Admin: Admin Indicators	
Admin: Financial Data	High level financial information, not at the detailed budget level, but in terms of bottom line and n-1 budget, debts, assets, etc
Financial: Budget	Detailed line-by-line budget breakdown
Financial: School Fees	Detailed student-by-student fees paid to the school for tuition, books, meals etc
Financial: Supply and Inventory	Detailed breakdown of available supplies including consumables, IT equipment, building materials etc
HR: Salaries	
HR: Employee Profiles	Teacher and staff contact information, educational background, licensing, etc
HR: Professional Development	Teacher and staff professional development (ex OKR/KPI) measures
HR: Cert and Training	Teacher and staff certification and training status, outcomes, scores, etc
HR: Disciplinary Records	Teacher and staff disciplinary records and outcomes
Outcomes: Grades	Individual student class-by-class grades
Outcomes: National Assessments	Exam scores for MoE driven tests
Outcomes: Classroom Assessments	Exam scores for classroom and school driven tests

vii. Portal - enables local community members to come and use the internet for e-Government functions Data is entered as a Pandas Dataframe Series with the following inputs:

1. Mean Value
2. Minimum value (3 SD below mean) as a percentage discount from mean
3. Max value (3 SD above mean) as a percentage added to mean
4. Frequency with which this data is uploaded (times per year),
5. Unit of measure - the group of users across which the data is being measured

Voter Registration	Ability to register to vote, check status, and update records
ID Renewal	Ability to renew existing ID through secure means
Annual Taxes	Ability to pay taxes, assumed to be on an annual basis, including linking to any electronics funds accounts
Bill Payments	Ability to pay private and public bills (ex utilities, transport charges) via government linked payments accounts
Complaints and Reporting	
e-Petitions	

viii. Demographics - semi-localized information (country level). These assumptions can be re-cast into a GEOTIFF or other format when locale specific information exists.

Student Teacher Ratio	Average number of students in a classroom (**not** simple #student/#teacher)
Teacher Classroom Ratio	Ratio of number of teachers to number of classrooms. Gives an approximate ratio of teacher load vs. planning time
School Age Fraction	Fraction of community that is of school age, according to the local age pyramid
School Enrollment Fraction	Fraction of school age children actually enrolled in school
People per Household	Average size of household
Skilled Labor Cost	Average cost for skilled and technical labor, in USD/hr
Manual Labor Cost	Average cost for manual labor, in USD/hr

2. Load in technology and infrastructure data. The data takes on two different forms, for communications hardware and energy hardware, represented in JSON lists. The inputs are:

    i. Comms technology

Speed	Max bandwidth of hardware as installed, MBps
Overnight Hardware - Fixed	Cost of hardware required to begin operations, not including financing cost (as if it was installed “overnight”), fixed costs for discrete hardware inputs such as modems
Overnight Labor - Fixed	Hours to install fixed hardware
Overnight Hardware - Variable	Cost of hardware required to begin operations which varies based on some parameter. For example, fiber costs which have a per-distance component
Overnight Labor - Variable	Hours to install one variable hardware unit
Setup Fees	Upfront fees to begin connectivity (ie ISP origination cost), in USD
Annual Hardware	Annual maintenance and replacement costs for hardware, as a performance of upfront fixed costs
Annual Labor Time	Number of hours to maintain hardware
Annual Fees	Annual ISP fees, in USD
Power	Electric power needed to run equipment, in watts

ii. Energy technology. Skilled labor, such as technicians, are charged a different rate than regular labor who would be responsible for digging, trenching, etc.

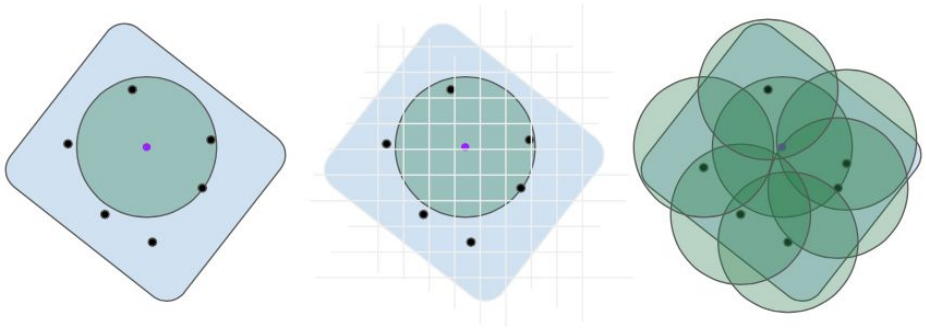
Overnight Hardware	USD per kW or kWh to purchase all components
Overnight Labor - Regular	Hours of regular labor per kW or kWh to setup system
Overnight Labor - Skilled	Hours of skilled or technician labor per kW or kWh to setup system
Annual Hardware	Cost to maintain hardware in USD per kW or kWh
Annual Labor - Regular	Number of hours (per kW or kWh) of regular labor required to maintain energy subsystem
Annual Labor - Skilled	Number of hours (per kW or kWh) of skilled or technical labor required to maintain energy subsystem
Daylight	Hours of usable daylight (for solar panels only). This number is then reduced down for productive daylight

	hours.

3. Compute school and community bandwidth requirements
    a. Compute School Bandwidth Requirements
        i. Compute School Census.

        If the number of students, teachers, and classrooms in a school is available from the ministry of education or other data source, this step is skipped and we move on to computing the bandwidth. If the number of students are known, we compute the number of teachers and classrooms from the project inputs described above.

        The school census is performed in computeSchoolCensus.py



1. Find the number of schools within a 10 km radius of the target school - the analysis assumes that the average student will travel up to 10 km to use the internet available at the school.
2. Compute total population inside the 10 km radius, using the UN adjusted 2020 100 m^2 dataset. Divide population among the number of schools, to avoid double counting. Uses the rasterio package to handle the GeoTIFF file.

    Note that this may overcount in urban areas, as students will attend schools closer in. Similarly, the algorithm may undercount in rural areas as students may be willing to travel further than 10 km especially if they have family support or encouragement to pursue education.

    Schools slightly less than 10 km from a cluster of schools will likely face some undercounting.

    This algorithm should be updated to only distribute students that lie between schools (assign each student to the ‘nearest’ or ‘neighborhood’ school).

3. Repeat the process for each school in the nationwide dataset. Assumes the dataset is both complete and does not have duplicates.

   Note that if schools are duplicated, the school census will still be reasonably accurate because each duplicate entry will have a portion of the student population. However, computed costs will be higher as each entry will receive its own connection.

ii. Compute School Bandwidth. This is performed in computeBandwidth.py The following use cases are modeled:

Use	Primary User	Use Category	Priority - for school
EMIS filings - Compliance, Student, and School Status	School Admins	Reporting	P0 - Critical for every school
Homework	Students	Instruction	P1 - Important for (almost) every school
Recorded educational clips	Classroom	Instruction	P1 - Important for (almost) every school
Lesson planning	Teachers	Instruction	P1 - Important for (almost) every school
Telemedicine consultations	Classroom	Healthcare	P1 - Important for (almost) every school
Health notices	Government	Healthcare	P1 - Important for (almost) every school

For each, the inputs from projectInputs.py are imported, then computed up to a kbps number based on the number of users shown in the primary user table.

b. Compute community bandwidth requirements
   i. Compute community census - using the same “computeSchoolCensus.py” script, but reducing the radius to 1 km as we assume that community members are much less likely to travel to use the internet.
   ii. Compute community bandwidth. This is performed in computeBandwidth.py. The following use cases are modeled:

Use	Primary User	Use Category	Priority - for school

Government e-portal access	Community	Government Services	P0 - Critical for every school
On premise internet access	Community	ISP	P1 - Important for (almost) every school
Telemedicine consultations	Community	Healthcare	P1 - Important for (almost) every school

4. Select technology

   There are a few general types of communications technology available to select from:

   - Fiber - direct connection from the nearest backbone directly to the school.
   - WISP - Wireless Internet Service Provider. This initial model assumes a line of sight microwave link and prices accordingly, but other technologies such as a WiFi mesh (for example) may also work well depending on the local context including geography and population.
   - Cellular - 2G, 3G, 4G. The model relies on accurate coverage maps, and does not assume any localized performance derating as this would require performing a networking connectivity survey at each location. Instead we derate the maximum link. If coverage maps don't exist, cellular is not selected.
   - Satellite - freestanding satellite mode.

   In reality there will be several different selection options from each hardware category, including vendors, models, etc. The model will continue to be updated as relevant quotes are received and initial deployments are made.

   Technology is selected first based on distance from fiber - all schools within 10 km of a backhaul will be connected. Next, all schools within 30 km of fiber will get connected to WISP. These are independent of bandwidth. Usage tends to grow to fill the available bandwidth, so maximizing the availability of high bandwidth connection is seen as the primary driver.

   For schools greater than 30 km from a fiber backhaul, tech is selected based on a combination of bandwidth and availability. We pick the fastest available cellular connection that meets the school's bandwidth requirements. If there is no sufficient cell connection, satellite is chosen last.

   In addition, power needs are computed based on the selected technology using the power requirements and generation/storage capacities in tech inputs, and the existing power infrastructure at each school site. The model computes:

   - Cost of solar - overnight and operational

- Cost of batteries - overnight and operational

Not every school will get the same power hardware. What gets installed will depend on the data provided. As a result, costs can vary dramatically between sites since the power needs are driven both by grid/generator and bandwidth needs. We assume that:

- Schools with no power data, or no grid connection and on-site hardware, will get both batteries and solar
- Schools with a grid connection and no generator or battery will get a battery
- Schools with a grid connection and generator don't need any power hardware
- Schools without a grid connection and with a generator will get solar

5. Compute Costs

Costs are computed in the computeCosts.py file. There are two main cost categories:

- Overnight Cost: the total labor and hardware cost to bring a school from its current state to being able to connect to the internet from there for the first time. "Overnight" refers to the assumption that the entire installation happens instantaneously - cost of capital and installation timelines aren't yet considered here since they will depend heavily on the .
- Annual cost: the average labor and hardware cost required on an annual basis to keep the setup running. These are not going to match up exactly to real annual costs, but instead should be thought about as an accrual. A device with a 5 year lifetime will have some non-zero probability of failing at any given time. The annual costs should be accrued with funds held in reserve, as the cost to replace a modem at any given moment is not the average annual cost but the total hardware cost. On the other hand, it's unlikely that *every* modem will fail at the same time, so accruing the annual cost for every school will net enough reserves to deal with the failures in that given year.