# AI Readiness Diagnostic Findings

## Step 4: Foundational Recommendations


## Step 4 Overview

*To find out how you can benefit from exploring your institution from an AI and Data Science lens, contact hello@educateventures.com*

- It is not always the best idea to select data sources that are **easiest** to access. Many times, it is worth going to a little more trouble to access data sources that are of the greatest **relevance** to our challenge
- Once we have identified works of relevance and we have identified **relevant data sources,** the next step is to **synthesise** all this data and information together. And don't forget, it is not just the data that you yourself might have collected or that from others in your organisation, there are sources out there that are **publicly available** and that can contribute to your understanding
- A few **initial questions** that need to be addressed for new data collection could be:
- Who is going to have **responsibility** for collecting the data? Perhaps it will be a course or module leader, or a class teacher, a head of the department, or a teaching and learning policy lead
- Timeframes: **when** will the data be collected? Today, next week, next month, next year?
- Over what **period of time** will the data collection happen
- And it is important to consider that this data collection could be occurring whilst teachers might be trying to keep up with their **schedules**
- Data can be collected by some common means, such as:
  - **Surveys**
  - **Interviews**
- But there is also **multimodal data**, such as video-enabled platforms, which can allow **analysis and reflection** of language used in classroom organisation or the addressing of **misconceptions**, or the style of **scaffolding** provided to individual students, or student-to-teacher **conversation**

• Key Takeaway:

• Data collection needs to be designed carefully and must **complement** that data which is already available. **Thoroughly examining your challenge** in the earlier steps should allow you to frame exactly what data you **do and don't want,** so that you're not wasting your precious time or opening yourself up to **risk**, collecting something that will later turn out to be irrelevant

# Recommendation: How to get started with data collection

*SUMMARY: don't just consider the data you've collected yourself, there are sources of data that you can access that are publicly available*

• It's time to explore how we can get started with data **collection.** We know that we can learn from **others** as we think about what's relevant to our **challenge** and **synthesise** it all (for more on challenges, visit **Step 2** in the AI Readiness Framework)
• Don't forget that it's not just the data that you might have collected or that from others in your organisation, there are some data sources out there that may be relevant to you that are **publicly available,** and that could contribute to your understanding of a particular challenge. **A few sources that you might find useful are below**
• Other **accessible evidence** that everybody can learn from is out there too, which can help you make a decision about what data you should try to **access,** and what data you should try and **collect**
• In **Step 3** of the AI Readiness Framework, the discussion is around the enormous variety of different data sources that could be available to tell us many of the things that are happening in terms of **behaviour** and **context** around our complex educational challenges that we're trying to unpack
• Think very **broadly** about data that you can access, that already exists, and when it comes to data collection, make sure you think about how **relevant** it would be to collect data about these particular **people**, or the particular **physical environment,** the **virtual environment,** information on the **resources** in use
• Don't forget that if there's data about connections **between** those different data sources, and connections between their data sources and the **people** that we're interested in, then it's really important to collect that too
• A final point is that **credible** data is really important: **accuracy** is paramount. As we move forward with collecting data, we will think about making sure the data we collect is accurate and that we collect it in as **organised** and as **precise** manner as possible

## Publicly Available Research Resources

• **Open Knowledge Maps**
The world's largest visual search engine for scientific knowledge. The engine dramatically increases the visibility of research findings for science and society alike

• **Digital Promise Research Map**
This map connects education practitioners to relevant and accessible learning sciences research findings that can help them shape powerful learning solutions for all students

• **Institute for Education Sciences: What Works Clearinghouse**
A site to select topics to find out what works, based on the evidence

• **John Hopkins University: Best Evidence Encyclopedia**
This website aims to empower educators with evidence on proven programs

• **Education Endowment Foundation: Teaching & Learning Toolkit**
An accessible summary of education evidence

• **Educause Research Publications**
This site conducts and distributes high quality research and analyses designed to identify trends, foster dialogue, and enhance decision-making

• **Google Scholar**
A freely accessible web search engine that indexes the full text or metadata of scholarly literature across an array of publishing formats and disciplines

# Recommendation: What new data should you have?

*SUMMARY: spend time planning what you want to collect, as whatever you find out about your challenge or concerns is only as good as the data you have*

• First, as we are looking to understand our specific challenge (that we decided on in Step 2 of the AI Readiness Framework), we need to recognise that whatever we **discover** about that challenge is only as **good** as the data that we have at our **disposal**

• If you haven't planned **appropriately,** and you end up collecting the **wrong data,** then you've wasted **time** and **resources.** It is really worth spending time **planning**

**• Who's responsible for gathering data?**

**• When will the data be gathered?**

**What is the timeframe for collection? In other words, over what period can you collect this data?**

**• You might also have data that's historically going back five years from the data that already exists, but what is your timeframe for collecting any new data that you're going to have access to?**

**• How will the data be gathered, transferred and stored?**

**• Is the data required to be in a particular format?**

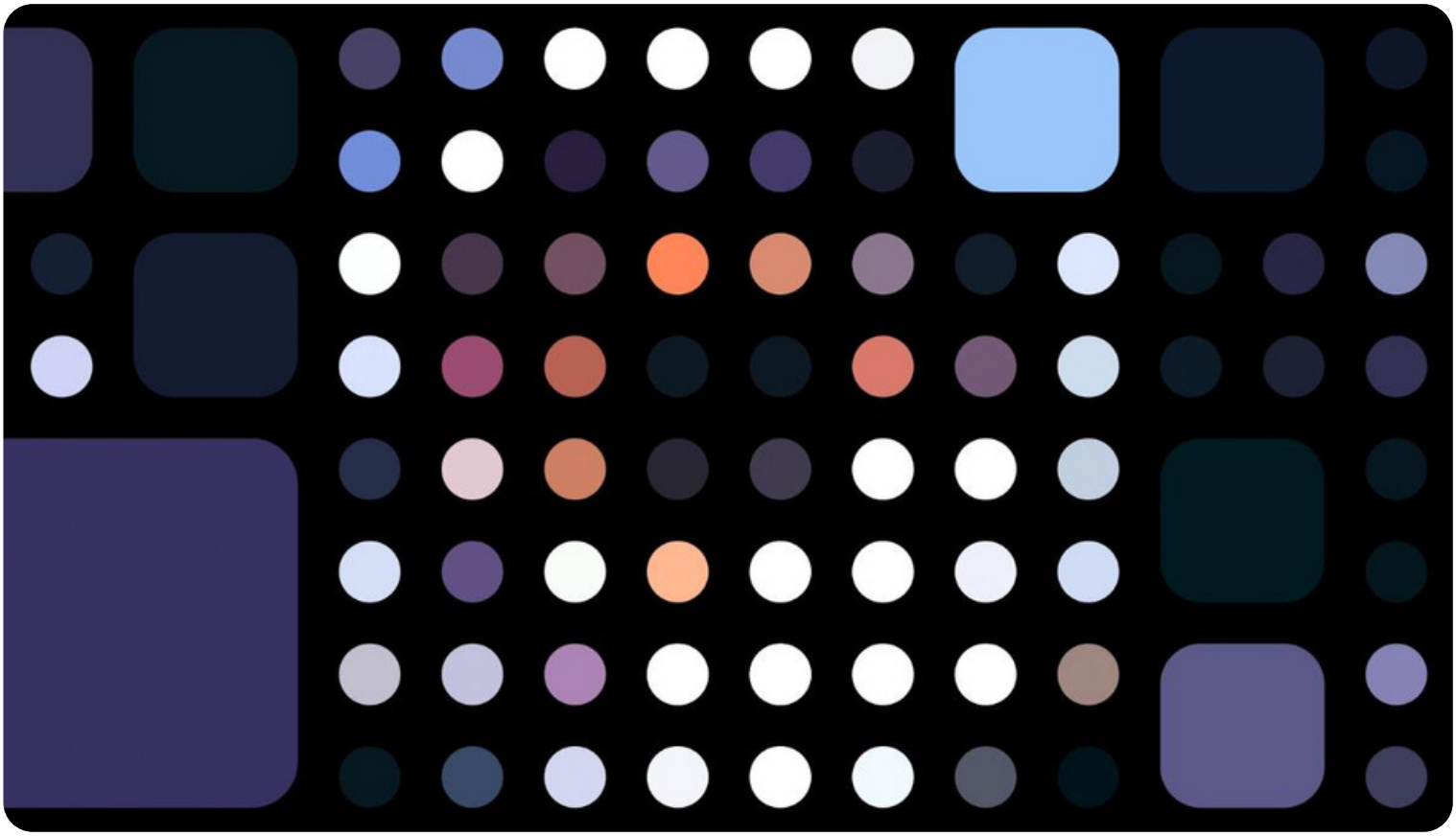
**Ethical issues must never be forgotten. If we collect data, do we have permission from the people whose data it is to collect it to use it?**

**• Are there any gaps in the data?**

• These are key considerations that we must explore. Ask yourself these questions when you're planning to collect data

• Considering our example challenge of trying to maintain the standards of teaching and learning online during the Covid pandemic: perhaps we already have data about the **identity** of all the students and all the educators and all of the **lessons,** the **sessions,** and the **activities** in which they were involved. We may already have some assessments, so we may have some **outcome data.** We might also have existing information about the type of **technology** that was being used - a laptop or a phone, for instance. We may have access to average **broadband** speeds, but it's not an easily accessible public resource everywhere

• We probably have a **timetable,** and we may have **video and audio recordings** of all the online sessions. So let's say that we have all that, plus any **chat transcripts,** but no recordings of small group or **breakout sessions.** That would give us a lot of data already. What are we going to collect to **complement** that dataset and tell us something that we can't get from accessing those existing data sources?



# Recommendation: Methods for data capture

*SUMMARY: although surveys yield self-reported data from individuals, they are exceptionally useful for gathering structured and unstructured data*

It is time to introduce **surveys** in the context of our challenge: that of trying to ensure the quality of teaching and learning is maintained online during the pandemic

A survey might help you to understand what's **happened** in a situation and how people **feel** about what's happened, for example. It's true that it will only be what that person **reports** to you that you **receive**, but perhaps it would be useful to know things like:

**The confidence people have about how much they felt they have been able to learn online - was doing it online more effective or less?**

**If they are an educator, what do they feel about the quality of what they've been able to do? Has it been as high quality as they would like?**

**Do they feel they're prepared for the differences in the way interactions are happening?**

There's lots of different things that we can find out through a survey, but how we can go about collecting **data** using a survey? The definition of a survey is: **information gathered by asking a range of individuals the same questions related to their characteristics, attributes, how they live, or their opinions**

Surveys have many advantages. They're **cost-effective** and can reach a lot of people online, and online surveys can be **anonymous,** for instance, which is useful for the **survey-taker.** If you think about the very simple forms of surveys that you might complete when you are going through the airport, you could be asked how you felt about the security process. The answers are happy, sad, unsure. That would be a very simple survey but you would get lots of data that way. You could say that the survey results represent **X% of an entire population,** if you have a large enough **sample** and a very carefully **designed** and **conducted** survey, rather than just **X% of survey respondents.** Surveys can be considered **reliable,** if questions are well **crafted** and a lot of care is taken to minimise **bias.** The responses can be **trusted**

You can have **closed-ended or open-ended questions.** With your open-ended questions, they are more complex to **analyse,** but it means you can get different **answers** from different people. They're often slightly longer than the **multi-choice survey questions.** But, you get more **information** and **explanations** that way

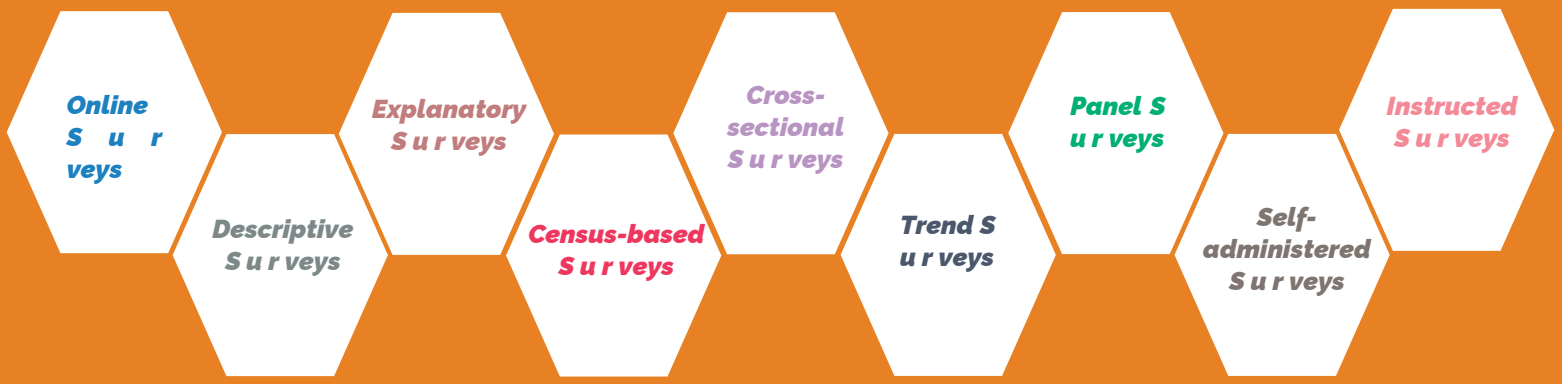
Surveying is a **specialist activity.** When it comes to making **population claims,** it is unlikely, for instance, that any **school** or **educational business** would have the **number** of participants and the **rigour** that would enable a claim to be made about the **entire population,** but nevertheless it is one of the potential **advantages** of a survey

When it comes to specialist activity, however, it's useful to understand there are lots of different **types** of survey. Some are descriptive, some are explanatory: we can use surveys to **describe** a particular population, or we can try and explain what's happened to the **behaviour** of a particular population We can conduct surveys that are census-based, or cross-sectional based. If it's a census-based survey, it's looking at **everyone.** Census is a **sample** including everyone. If it is cross-sectional, then you're looking at a **portion** of the overall population, maybe all men or women, or all teenagers. When you think about the census that's done in the UK, that census is about everybody in the population, and a huge amount of effort is put in to try and make sure that everybody answers it

You can also have trend surveys or panel surveys. **Trends** are looking for changes over **time,** and actually, we may think about the number of **cycles** that are needed in order to claim that trend. You're going to need to think about how to get **trend data,** which you can't get from just **one** sample. You need to think about the **number of cycles** that you will have to put in place, in order to get that trend data, and how frequently you need to **repeat** the survey. If only **two** data collection points are possible and you perhaps see some sort of increase in a **variable,** you can't really say it's a **trend**

For example, maybe you conducted a survey in **January,** and it's about how much people drink, and people are drinking 10 units a week on average. You conduct the same survey in **June** and people are drinking twenty alcoholic units a week on average in June. You can't say there's a trend for alcohol to **increase,** because you have only got **two** data points. It could be an **anomaly,** particularly in that instance, because many people do 'Dry January', and don't drink that month
A panel survey is just a **snapshot.** So you've got the option to collect surveys **repeatedly over time** if you want to identify trends. But you can also just have a single snapshot, which we call a panel survey

And you can have surveys that are self-administered. They are sent to people, maybe online or in the post, or you can have a survey where there's a person who does it with you, this is an instructed survey where there's more care taken to make sure that people **understand** the questions and they can ask for **clarification**



# Who can help me?

*We are specialists in ethical AI solutions for schools and education and training businesses - contact our team for help*

The EDUCATE AI and Data Science team was formed to consult on and co-design ethical AI solutions to complex problems in data-driven technology ventures and schools. Our team of computer scientists, educationalists, and world-renowned experts can take you from zero AI to a comprehensive evidence- led strategy and beyond, with effective, scalable AI-powered teaching and learning solutions.

To find out how you can benefit from examining your institution through a **'data and AI lens'**, and leveraging the transformational power of AI to tackle your challenges, contact the **AI and Data Science Team** at EDUCATE Ventures Research at **hello@educateventures.com**.

Thanks for reading!

**- The EDUCATE Ventures Research Team, 2023**

# Further Reading

*Below you can find a selection of resources, books, podcasts, webinars, and research papers appropriate to your stage of AI Readiness. Good luck!*

- AI for School Teachers, Byte-Sized Edition

An easy-to-read 10-page byte-sized summary of the book of the same name, written by Professors Rose Luckin, Mutlu Cukurova, and Headteacher Karine George, members of the senior team actively developing and using the AI Readiness Framework from which these recommendations derive

- Hey Mum, the Fridge has just let the Burglars in

A short blog from the Institute of Education featuring the dangers of unethical data capture in devices with poor online security

- AI Readiness: Step 4 webinar for Educators/Businesses

Two separate webinars introducing Step 4 of the AI Readiness Framework, one targeted toward educationalists, and the other targeted to educational businesses