|  |  |
| --- | --- |
| **Slide 1** | Greetings Everyone. My name is Ian Hambleton – I’m Professor of Statistics at UWI. I’m leading a project known as CaribData.  So, for almost 30 years now I’ve been collecting and using Caribbean data to generate evidence for the region. And this project in many ways builds on the experiences I’ve had – and I hope and suspect those experiences will resonate with you all as well. |
| **Slide 2** | So I’m going to spend a few minutes expanding on WHAT we’re planning for CaribData, and WHY we think this project is needed and is good for the region.  And these are the 4 themes running through our project.   * Infrastructure. * Training & Guidelines * Analytics & Communication. * And Collaborative working |
| **Slide 3** | Broadly, using existing UWI infrastructure as our foundation - we’ll be building a set of resources to enable and encourage Caribbean data sharing and data re-use  And sticking to the same four themes:   * **NEXT.** We’ll be setting up an online infrastructure to enable easier data collection, data sharing and re-use, * **NEXT.** We’ll be developing training programs and longer-term mentoring to help increase capacity in data handling and data analytics – and we’ll be co-creating these programmes with our beneficiaries. * **NEXT.** We’ll be performing centralized analytics along with an active journalism-driven communication program to showcase existing data, champion data re-use, and to actively communicate the stories around the data we all produce * **NEXT.** And through workshops and other public events, we want to start building more collaborations that could lead to more data use across disciplines – more data linkages. For example, we have funds for an annual data-storytelling datathon.   So we have round one funding from the IDB, and our official project beneficiaries are NSOs. |
| **Slide 4** | Why do we think our project can be important for the region.  So the data industry is moving fast.  Some of us can remember when handling of data used to be a simple affair. Collect it, clean it, use it, store it.  The move towards handling ‘big data’ – so large volumes and varieties of data, collected quickly – is one reason why data handling is becoming more complicated. Data collection is now mostly online. It is sometimes automated, and the associated data cleaning requires us to understand different subject matters and specialist software.  Just as important, data handling infrastructures must now conform to a fast-changing set of regulations, and these define for example minimum standards of security and participant confidentiality.  And countries across the Caribbean are updating their Data Protection Laws.  **So one goal of our project is to help Caribbean data professionals keep up with best practice data handling.** |
| **Slide 5** | I’m sure that each of us at some point have received datasets that are less than perfect.  And it’s not too dramatic to say that data prepared without best-practice data handling is regularly hard to use, time-consuming to correct, and sometimes unusable.  For data to be useful in the long term it must be accurate, and the data must be well documented and easily available.  Data that don’t meet these basic criteria are stored at institutions across the world, creating what we might call ‘data graveyards’. And these graveyards use resources that would be better spent elsewhere.  **So another goal of our project is to reduce data wastage across the Caribbean. Again – this is all about making data accessible.** |
| **Slide 6** | Across the Caribbean, international collaborations are common, and the data that come from these collaborations are often collected and stored using overseas data infrastructures. This is great if it creates high quality data and stores that data securely.  Often though, we hear of examples where data are hosted overseas without a full understanding of long-term data access or the implications for looking after that data. And we see this as a very important risk in the creation of long-term data resources for the region.  **And so a goal of our project is to offer a local best-practice data infrastructure to give the region greater control of its data resources – we want to encourage data-sovereignty for the longer term.**  Now, we will never force this infrastructure on those that don’t need it – but it will be a robust regional option for those who are interested. |
| **Slide 7** | Last and perhaps most important for our project, data availability and accessibility in the Caribbean remains limited  There are lots of examples of this problem, and here is our analysis that’s been accepted for publication later this year.  The analysis uses World Bank Statistical Performance Indicator Framework (2022). |
| **Slide 8** | The framework covers many aspects of national data readiness, and I’m looking here at data availability.  First for the SDGs.  And here is a graph of the 7 world regions, and at the bottom we see that between 2018 and 2022, SDG data availability for the SIDS was lower than for any other world region, roughly equivalent to availability in sub-Saharan Africa (SSA), despite very different development profiles.  Now I reckon that the SDGs probably represent a best-case scenario for data availability, given the level of support globally for the SDGs. So let’s now look at data availability across 900 gender-stratified metrics.  ---  EXTRA  Two-thirds of SDG indicators were available for SIDS in 2022; other world regions having between 71% and 87% availability. |
| **Slide 9** | And again the SIDS are lower than for any world region, and this time the shortfall is more substantial.  One-third of gender-stratified indicators were available for SIDS in 2022; other world regions having roughly two-thirds and higher availability. |
| **Slide 11** | And perhaps unsurprisingly, population size is strongly associated with data availability,  Here is a graph showing the relationship between national statistical capacity and population size. The SIDS are in orange, all low income nations are in blue. By and large, the SIDS have lower statistical capacity than other low-income countries  And this pattern the same irrespective of World Bank income group.  So a lot of evidence that the absolute resources available to SIDS affects their data capacity.  --------  EXTRA On average:   * with SIDS having a statistical capacity 18 percentage points lower than other countries, * SDG availability 17 percentage points lower * gender-stratified indicator availability 29 percentage points lower |
| **Slide 12** | So our projects has 6 sub-projects:  **1st**, we will be contributing to scientific publication on the theme of open data, and I’ll come back to this in a moment.  **2nd**, I said at the start that the project is about enabling and encouraging data sharing. Part 2 is the enabling part, and we’ll be hosting platforms for data collection and data sharing. Our data collection platform is already operating, and we’re running an instance of REDCap in a GDPR compliant environment.  **3rd**, part 3 is all about encouraging data sharing, and for me, this is the core of the project. We have a journalist on our team, we’re building a community of interested data communicators, and we’re developing a website for hosting hopefully engaging data stories using openly available data from the region. So part 5 is all about data communication.  **4th**, Part 4 is about training. We are planning several short courses on aspects of data sharing and data communication. More importantly I think, we planning longer-term mentoring for early career data scientists. For example will we’re taking Data Science graduates for 6-month placements as part of the ICAMP program – Intra-Caribbean Academic Mobility Programme.  **5th**, there is a lot of regional commentary about the lack of available Caribbean data, and much of this dialogue is informal – without a clear evidence base. So we’ll be developing methods for systematic reviews of regional data availability, which is not trivial, and is similar to a grey literature SR.  **6th**, This project will have failed if it can’t continue beyond its initial funding. IDB are seeing our work as a pilot for future funding. But we don’t want to rely on IDB, and we are partnering with business to develop a pragmatic sustainability plan. |
| **Slide 13** | We’re well into this first sub-project, partly to make sure we publish during the life of the funding.  And here are our two planned flagship publications:  The first is a 100+ page analysis of the WHO GHE in collaboration with PAHO. This has just been released, and we’re currently converting the PDF publication into a web publication.  And later this year we’re partnering again with PAHO to produce a special edition of the PAJPH on “Open Data for Caribbean Health”. And with colleagues from across the region we’re producing 13 articles on a wide range of topics relevant to open data. This is slated for release late 2024 / early 2025. |
| **Slide 14** | Here are our current partners, and we are actively looking to expand our CaribData partnerships. |