7th International Conference on Big Data and Data Science for Official Statistics, Yogyakarta, 7-11 Nov 2023

Prepared by UN Committee of Experts on Big Data and Data Science for Official Statistics

Table of Contents

Introduction

The UN Committee of Experts on Big Data and Data Science for Official Statistics (UNCEBD) was established in 2014 to provide strategic vision, direction and coordination for a global programme on big data for official statistics, including for indicators of the sustainable development agenda. The Committee was tasked also to promote capacity-building, training and sharing of experience and to foster communication and advocacy of the use of big data for policy applications. One of the ways to achieve the outreach goal was by conducting an international conference on big data and data science, at which not only statistical institutes were brought together, but also other government institutes, private sector organizations, academia and civil society. The first conference was held in October 2014 in Beijing[1], China; followed by conferences in Abu Dhabi[2], UAE, in 2015; in Dublin[3], Ireland, in 2016; in Bogota[4], Colombia, in 2017; and in May 2019 in Kigali[5], Rwanda. Due to the global health pandemic, the conference in 2020 was held virtually organized with Statistics Korea[6].

Political support and partnerships with private sector and academia are vitally important to successfully use big data and data science in the production of timely, frequent and granular statistics and indicators. The first day of the conference has, therefore, consistently been used to draw high-level political attention, such as the Prime Minister of Rwanda speaking at the 2019 conference. The themes have changed over the years, such as the theme of "Data collaboratives and trusted data" in 2017, and ""How can Big Data help in the COVID-19 response?" as the main theme in 2020.

In November 2022, the statistical community and its partners could be brough together again in person. The national statistical office of Indonesia (BPS Indonesia) was willing to host this event. BPS Indonesia has been one of the leaders in the South-East Asia region of the use of big data and data science, especially in the use of mobile phone data for tourism and dynamic population statistics. The timing of the conference coincided with the presidency of Indonesia in the G-20, which had the theme "Recover together, recover stronger". In the same spirit "Global Challenges and the Importance of Relevant and Timely Data" was chosen as the overarching theme of the 7th international conference on big data and data science for official statistics.

All of the information of the conference can be found on the website of UNCEBD at https://unstats.un.org/bigdata/events/2022/conference/agenda. The information covers the integral recording of the first two days of the conference, as well as recordings of the workshops which took place on the third, fourth and fifth day of the conference. Furthermore, you can find the presentations of the keynote speakers on Day 1, and of several speakers on Day 2. Besides the recordings of the workshops, you can also find workshop materials for downloading, see https://unstats.un.org/bigdata/events/2022/conference/workshops/index.cshtml

---

[1] http://unstats.un.org/unsd/trade/events/2014/Beijing/default.asp
[2] http://unstats.un.org/unsd/trade/events/2015/abudhabi/default.asp
[3] http://unstats.un.org/unsd/bigdata/conferences/2016/default.asp
[4] https://unstats.un.org/unsd/bigdata/conferences/2017/default.asp
[5] https://unstats.un.org/unsd/bigdata/conferences/2019/default.asp
[6] https://unstats.un.org/unsd/bigdata/conferences/2020/

Theme of the 7th International Conference: Global Challenges and the Importance of Relevant and Timely Data

The global economic recovery momentum from the impact of the COVID-19 pandemic has been disrupted due to the ongoing geopolitical uncertainties. Economic slowdown, rising inflation and the associated increases in interest rates have dampened near-term economic and sustainable development prospects and pose additional fiscal and debt challenges to an already constrained fiscal space. At the same time, the human and economic costs due to the ravages of climate change continue to rise, disproportionately affecting the poor. To navigate through these difficult circumstances and to take appropriate policy actions, governments have to understand and monitor the existing situation and therefore need relevant and timely data. The statistical community is tasked to deliver these data, including data for an effective pursuit of the Sustainable Development Goals.

Thorough assessment of emerging development issues and challenges require timely and flexible production of indicators often in collaboration with strategic partners from public and private sector as well as academia. The first day of this Conference was meant to engage with the policy makers and other stakeholders, which have the data demand. In the context of a confluence of shocks (COVID-19, conflict and climate change), sustained economic recovery is a major theme among policymakers.

The program on the first day will consist of high-level dialogues around the main themes of "Global Economic Recovery", "Food Security" and "Access to Relevant Data". The second day will have dialogues with experts of the global statistical community. The United Nations Committee of Experts on Big Data and Data Science for Official Statistics will review in an open session the work of the past year and, more importantly, will discuss the way forward on incorporating big data and data science into the standard business processes of statistical organizations. The latter part of week will showcase the 2022 Big Data Hackathon as well as several workshops.

Global Economic Recovery

Rising energy prices are accelerating the cost-of-living crisis and sustaining the vicious cycle of constrained household budgets, food insecurity, energy poverty, and rising social unrest. The crisis is deeply impacting vulnerable populations in developing countries, threatening hard-won gains in the access to energy.

Food Security

Supporting the livelihoods of small-scale food producers, improving the resilience of food production systems and encouraging the sustainable use of natural resources are all key to fulfilling this mandate and achieving Sustainable Development Goal 2 (SDG2), a world without hunger, food insecurity and malnutrition.

Access to Relevant Data

The Conference of European Statisticians in October 2022 called " upon national governments and relevant ministries to support national statistical offices in their key coordinating role in the access to data held by the private sector for official statistical purposes, bearing in mind the critical role of high-quality, independent and impartial official statistics..."

Day 1: Dialogue with Policy makers

Opening Ceremony

Mr. Ashwell Jenneker, Deputy Statistician General, South Africa and Acting Chair of UNCEBD

The committee of experts on big data and data science for official statistics was given the mandate in 2014 to explore the benefits and the challenges of using the available digital data for the compilation of official statistics. Digital Data is everywhere around us. On our mobile phones, on satellite imagery, when we make payment transactions or keep track of how many steps I've taken this morning. That is how digital technology is changing the world. We can immediately see what's happening. Now the private sector has been using this for a while; so, they are ahead of us. Digital data will allow us to give results at a much more local level and more frequently. If we do a census, we do this only every 10 years. With big data we can get detailed information more frequently.

The conference is organized at a time where we are all suffering from tremendous economic hardship. For example, in my country the economy was struggling even before Covid-19, and Covid made things worse. There's food insecurity, there's climate change, there are wars in some parts of the world. These are the challenges that this conference is faced with; and we need to find new ways to inform the decision makers; so that they can make appropriate policy decisions.

Indonesia is hosting the G20 at this moment and we as statisticians need to make sure that we are represented at the table; and we can only be represented at the table if we've got timely and relevant data. So, we need to make sure that we embrace new technology, new data and new ways of making data available. Budgets get cut around the world. Also in South Africa budgets are cut; and unless our chief statisticians are sitting at the right tables, you might find that our budgets get cut more and more.

There are also other reasons why this conference is special. I mean that we've got almost 2000 young statisticians who are going to take part in our hackathon. The hackathon will look at current day problems and see how we use data and technology to resolve that. We haven't had that before, where we've had young statisticians taking part right across the globe in our conference. I mean that transforms the way we look at things, because you've got these young creative minds looking at problems differently from the way we've been looking at it because when you are an official statistician you've got a certain framework and a certain rubric within which you look at things. Young people look at things differently.

In South Africa this year for example, we had the census. It takes us 3.2 billion Rand to do the census this time around because for many different reasons people didn't necessarily want to open their doors. So, we had to extend the census three times to finish it, resulting in a huge overspending. As time goes by it will become more and more difficult for us to conduct official statistics in the traditional way. Now in South Africa we've got more cell phones than people. A cell phone always projects where you are. So, the question is can you use cell phones to count the number of people in South Africa? You can look at nighttime and daytime populations. You can look at migration patterns. Imagine if you can count all the people in your country almost every month. There are of course challenges which we will put before the young statisticians to solve. This creates hope for the future.

Thank you.

Ms. Anjali Sen, Country Director UNFPA in Indonesia

I'm here to represent the UN team in Indonesia and our resident coordinator Ms. Valérie Julliand. This conference is indeed held at a very auspicious time, as was also mentioned by Ashwell. In the midst of growing global challenges when timely access to relevant data is particularly crucial to help tackle these global challenges of sustainability, economic recovery and food security, innovation in data science and big data is of utmost importance. I was very inspired to hear about the 2000 young statisticians who will indeed be coming up with innovative solutions and guiding us in this field.

While Covid-19 remains a threat to social and economic stability, the war in Ukraine and the resulting cost of living crisis are taking a toll on people's lives and livelihoods across the globe, and of course this means greater uncertainty for economic recovery. Economic condition hampering the progress towards the sustainable development goals on many fronts, and particularly amongst the poorest and the most marginalized communities; and this is where we really need to step up our efforts and really redouble our efforts to get the sustainable development goals back on track. While the world is bursting with information, there is still a lack of indicators, where they are most needed: data on the poorest and the most vulnerable and marginalized communities and if indeed we are serious about leaving no one behind, which is the central premise of agenda 2030, then we need to be serious about data. No SDG target can be met until it is met for all everywhere.

There is an urgent need for us to harness the power of data, as we confront these many global challenges. We know that with reliable data we can ensure that we can build back better from the pandemic and deal with this double crisis of energy and food insecurity. The data revolution encompasses the open data movement, the rise of crowdsourcing, the new ICT for data collection and an explosion in the availability of big data, which was also explained very eloquently by Ashwell, and these along with the emergence of artificial intelligence. The Internet of things can help accelerate the achievement of the SDGs really ensuring that no one is left behind.

The United Nations is committed to support the use of big data for policy making and sound policy decisions and response. With Indonesia as the host of this conference, I'm indeed very privileged to be here as a representative of UNFPA for the last three years. This is a beautiful country with wonderful people, and we have a strong partnership with the government.

The United Nations has supported the government in this endeavor to fill information gaps in the development and humanitarian sectors through the adoption of big data and application of the survey design with emphasis on gender equality, disability and social inclusion. The UN also supported the Covid-19 response through the development of various Big Data platforms and approaches. For example, we used innovative data collection systems to support government interventions for children and families that were affected by the Covid-19 pandemic, and we also supported the deployment of digital platforms to support vaccine distributions.

This year also marks the 10th anniversary of the Pulse Lab Jakarta which is a joint facility of the United Nations and the government of Indonesia. They have carried out numerous exploratory research activities using a broad range of Big Data sources and applying its analysis to support effective and inclusive development and humanitarian action. In this respect we will continue to support the government and the key stakeholders to harness these new and advanced Big Data practices and to leverage partnerships for the attainment of the SDGs.

Ms. Armida Salsiah Alisjahbana, Under Secretary-General and Executive Secretary of UNESCAP

This meeting is held at an important time where our world faces many global and local challenges across a wide range of topics including economic recovery, climate change, health and the environment. Coordinated action to address shared challenges and to meet the 2030 Agenda for Sustainable Development has become even more important over the last few years and new and emerging issues demand increasingly timely and flexible data.

ESCAP's SDG Progress report showed widening disparities in the region and COVID-19 has been a huge challenge for all countries in Asia Pacific and for us all as individuals and communities. Progress towards the SDGs in the Asia Pacific region has slowed as the COVID-19 pandemic and climate change have exacerbated development challenges. And the region is not on track to achieve any of the 17 SDGs.

As the statistical community, we can be proud of our work which makes it possible to not just measure the SDGs but more importantly to achieve them. I am pleased to see that the availability of data in Asia and the Pacific has sharply increased since the first SDG benchmark report in 2017 – indeed the number of indicators with sufficient data has almost doubled since then. This trend testifies to the tremendous effort which has been made to strengthen national statistical systems in the region to respond to the monitoring demands of the 2030 Agenda. However, there are still very significant data gaps; especially where traditional data are missing, Big Data can be part of the solution to help to monitor the Sustainable Development Goals (SDG) indicators.

Big Data already permeates many aspects of people's lives, from daily communication and interactions, to shopping and consumption and the medical treatments they receive. Big Data is also transforming the way people and businesses make decisions and measure things. With an invaluable continuous flow of digital information about activities and their impact on society, the economy, and the environment, Big Data holds tremendous potential for official statistics. I have been pleased to see the statistical community increasingly turning your attention to new data sources despite numerous legal, technological, and financial challenges.

As it is constantly being collected and generated, Big Data provides timely, frequent, and granular insights - crucial attributes in critical situations. These timely and detailed data can inform policy and help assess crises immediate impact on society and the economy.

I am hopeful this conference will provide an important opportunity to exchange the newest practices in countries and share knowledge of emerging methods and standards for using Big Data and data science for official statistics. I am sure your discussions on the way forward to incorporate Big Data and data science into the standard business processes of statistical organizations will provide an important steppingstone to faster data and statistics and therefore responses to the challenges we are currently facing and will be facing in the years ahead.

Ms. Amina Mohamed, UN Deputy Secretary-General

It is my pleasure to address you all this morning at the occasion of the opening of the 7th International Conference on Big Data and Data Science for Official Statistics. While you are all gathering in Yogyakarta today, only a few hundred miles to your East, the G20 economies are meeting in Bali to discuss the global economic recovery.

Recovery is a key concern at this moment, and requires concerted and well targeted efforts based on relevant and timely data, which is the well-chosen theme of your conference.

Our world is facing grave challenges. The Covid-19 and the war in Ukraine have disrupted supply chains and caused major spikes in energy and food prices, fueling a cost-of-living crisis for many millions around the world. Simply meeting the basic needs of food, water and shelter for their families is becoming an impossible struggle. At the same time the climate catastrophe is worsening, testing the resilience of our societies against typhoons, floodings and other severe weather events. To effectively combat these challenges more timely and relevant data to inform our policies and monitor progress will be essential. This is why you, the global data community, are so critical to our collective efforts. We rely on you to bridge data gaps and provide a clear understanding of where and how we can do better and deliver deeper insights and smarter decision making.

I'm greatly encouraged that almost 2,000 young data scientists and statisticians from every part of the world will compete in a data hackathon to find creative solutions to addressing the many challenges we face, above all climate change. The engagement of young people is crucial. We must ensure that the voices of those with the greatest stake in our future are heard loud and clear and I would like to thank the United Nations Major Group of Children and Youth for your enthusiasm and dedication.

I was also encouraged to learn that many national statistical offices, international agencies, academic institutions and major tech companies are supporting the UN youth group to make this hackathon a resounding success. Together we can meet the test of our time and build a better brighter future for all.

Mr. Imam Machdi, Deputy of Methodology and Statistical Information, Indonesia

First of all, on behalf of BPS-Statistics Indonesia, I would like to warmly welcome all of you to "the 7th International Conference on Big Data and Data Science for Official Statistics". This event is held by United Nations Statistics Division (UNSD) and BPS-Statistics Indonesia (BPS).

Previously, this conference is held by the UN Statistics Division in many different countries. In this year, Indonesia has the honor to host this prestigious event. It was not without reason that we decided to hold this event in this beautiful city of Yogyakarta. Jogja is a city with many predicates, among others: the city of tourism, city of culture, and most importantly, city of students or city of education. Its role in the development of education in Indonesia has been widely acknowledged. It is our sincerest hope that the love and passion of this city for education will spread to all the participants, which will contribute positively to the achievement of the meeting objectives.

In recent years, the world is facing numerous global challenges related to the environment, climate change, health, food insecurity, economic recovery, and human displacement. Governments must face the challenges by taking appropriate actions through relevant and timely data. The statistical community has been tasked with delivering these data, particularly data on the Sustainable Development Goals (SDGs).

This is why this conference is extremely important. It brings together many parties to address and discuss about the global challenges and the role of relevant and timely data to support evidence-based policy making. The theme for this conference is "Global Challenges and the Importance of Relevant and Timely Data". It covers three main topics: "Global Economic Recovery", "Food Security", and "Access to

Relevant Data". In addition to the conference, we also arrange workshop, hackathon, and exhibition. Through these activities, we hope that at the end of the conference, all participants will have awareness and eagerness to contribute to the provision of relevant and high-quality data as a basis for effective policy making.

Before I close my remarks, on behalf of BPS, I would like to express my deepest appreciation to all of the speakers for sharing your valuable knowledge. In particular, I would like to extend my gratitude to the Governor of Yogyakarta for the warm welcome and support for the success of this event; and to the Coordinating Ministry of Economic Affairs and the Ministry of Tourism and Creative Economy. Special thanks to UNSD for the excellent collaboration in organizing this event. Also, thanks to our sponsors: Telkomsel, University of Tokyo, DataProvider.com, Oblivious, Location Mind, UN in Indonesia, AWS, and Tourism Agency of Special Region of Yogyakarta.

Last but not least, I also would like to extend my appreciation to the committee and supporting institutions for the hard work, dedication, and valuable contribution to make this conference run smoothly.

I am confident that with firm determination, enthusiastic participation, and strong collaboration from all participants, this event will achieve great success and give positive contribution to development of statistics in the upcoming years. Finally, I wish you a fruitful conference and enjoy the beauty, warmth, and hospitality of Yogyakarta.

Mr. Sandiaga Salahuddin Uno, Minister of Tourism and Creative Economy, Indonesia

It is truly an honor for me to be able to share matters regarding data insights on global economic recovery to sustainability. This year, the accumulated number of travelers until August 2022 is around 1.73 million, increasing by 2000 percent compared to the same period in 2021. The growing travel demand, the change of preference, and decisions to travel have to be acknowledged and responded expeditiously by using big data. Through this conference, we expect it to become a medium for discussions on the utilization of big data as a generating tool that is relevant and punctual to support policy-making decision. With big data, we hope you can speed up the economic recovery process in Indonesia.

Mr. Airlangga Hartarto, Coordinating Minister of Economic Affairs, Indonesia

As a government institution responsible for providing data, BPS-Statistics Indonesia, is required to be independent so that the data produced by BPS can be used by stakeholders to produce targeted policies, including in Indonesia's economic recovery.

In the implementation of Indonesia's economic recovery strategy, BPS plays a role in providing strategic indicators that can be used by the government, in particular to see various phenomena and economic turmoil in Indonesia. Thus, the government will be able to formulate policies that support the economic recovery, such as economic growth, inflation, poverty, employment, tourism, agriculture, industry, exports, and others.

Given the crucial role of BPS in supporting Indonesia's economic recovery strategy, the provision of data that is fast, accurate and valid must be the concern of all BPS personnel. Accuracy is one of the most important criteria in evaluating the quality of statistical data. BPS in presenting economic data,

especially GDP, always tries and strives to provide GDP data with a lag of 35 days since the end of the first quarter, which is relatively faster compared to other countries.

On the other hand, to provide data on economic growth, basic data are needed both from internal and external. The complexity of the data needed to calculate the Indonesian economy has to be communicated and coordinated between work units and government agencies, so that the basic data is fast and accurate.

In providing data on economic growth, BPS also uses big data to enrich the existing data. In addition to GDP data, BPS is also able to release inflation data at monthly intervals, which is released after the month ends. In addition to inflation data, several data and indicators that are also released on a monthly basis are data on export import, hotel occupancy rates or tourism. The provision of fast and accurate data will support the effectiveness and efficiency of preparing economic indicators.

The era of big data creates many interesting opportunities for new developments in the economy (economic triggers). Big data enables more accurate predictions and improves causal inference. But on the other hand, big data analysis raises problems from methodological perspective that must be handled quickly.

Big data is able to capture people's consumption patterns in the digital zone. Consumption behavior is reflected through public consumption on e-commerce, public sentiment on social media, and the needs from their search patterns via the internet. Big data can also be used to view transactions that occur digitally, including complex transactions, such as household transactions, credit, company developments, export and import.

Provision of reliable and quality data and information is especially important to stimulate economic activity. Big data provides faster and more detailed statistical data. Subsequently, analysis on big data is certainly a necessity for business in today's digital era. Big data also has several challenges:

- Source of big data still depends on ethics and regulations.
- The statistical methodology must be able to guarantee the calculation of indicators which comes from big data.
- Technology, human resources, and infrastructure which are prerequisites for the use of big data must be enhanced.

The government and BPS are expected to be able to use big data to improve the quality of official statistics. BPS is expected to improve the frequency of data release, to provide estimates to smaller areas (Small Area Estimation), to reduce the effects of non-response, and to be able to provide information as a comparison. The use of big data is expected to improve the quality of from time to time. Thus, the national development achievements can be designed in a more inclusive manner in order to increase competitiveness for people's welfare.

This conference is an effort to enhance the capacity and capability of data science in Indonesia. Academics, practitioners and government officials are expected to use big data in new economic patterns and policymaking.

Sustainability of the global economic recovery

This first high-level panel discussed economic policies, monetary and fiscal policies, in particular, those that can help policymakers to not only steer economic recovery but also continue to pursue broader sustainable development ambitions. The panel focused on the importance of data for governments to support evidence-based policy making, as governments are faced with difficult policy choices.

Mr. Navid Hanif, UN Assistant Secretary-General for Economic Development

On behalf of the department of economic and social affairs, I would like to welcome all the participants of the seventh International Conference on Big Data and data science for official statistics. You are meeting against the backdrop of poly-crises that the world is facing today. These are multiple and diverse. The world is still reeling from the aftermath of the pandemic, and this is compounded by the war in Ukraine, inflationary pressures, monetary tightening, shrinking fiscal space and growing debt distress. We have seen increased in poverty and hunger. The ravages of climate change are exacerbating these challenges. The first panel discussion will focus on these difficult questions of how to bring the world on the path to sustainable inclusive and resilient recovery.

I would like to focus on four key messages for your consideration. First a message of hope: these challenges are surmountable. We have the solutions. We need to act urgently and with clear focus on people and the planet. That is where the role of timely and comprehensive data is critical.

Second, we never before had access to real time comprehensive data that policy makers can use to sharpen their policy choices. I find that extremely important in setting inflation targets and making monetary decisions. Similarly, data is critical for fighting inequality and providing social protection where it is needed. We have it evidence-based policy decision will help us to steer through these troubled waters.

Third, we have to rescue the sustainable development goals. Progress is being lost we have seen reversals and many SDGs are stalled. We need to bring them back on track for realization by 2030.

Finally, strong multilateral action cannot be overemphasized. No single country can solve the challenges of climate change inequality and growing social discontent. I would like to commend the G20 presidency of Indonesia for choosing "recover together recover stronger" as the overarching theme for the work of G20 this year, and I'm confident that the summit will send a strong message to the world for urgent and collective action.

I'm confident that my colleague Mr Hamza Malik will provide deep insight in the current challenges and suggest pathways for policy makers to address the short-term issues but also remain focused on the commitment to transform societies as agreed in the 2030 agenda, the Climate Action agenda and the Paris agreement.

Dr. Hamza Ali Malik, Director Macroeconomic Policy and Financing for Development Division, UNESCAP

Dr. Malik focused on recent crises and economic prospects. The two crises in particular he was referring to are the covid-19 pandemic and the war in Ukraine, and their implications for the economic prospects.

He was touching upon five key areas:

(1) The GDP Centric development approach which comes at a huge cost to social inclusiveness and environmental sustainability; so, there is a need to actually move away from this GDP Centric policy making and invest in people and the planet

(2) What's happening nowadays in the economies in the aftermath of the war in Ukraine and the pandemic; what kind of challenges are being faced.

(3) The policy constraints, which the global policy community is actually facing; fiscal positions have been constrained; debt is on the rise; inflation is high; interests are increasing. This is not the kind of situation um policy makers normally envy for or aspire for to be operating in.

(4) The need to move away from growth-centric models to more resilient inclusive and sustainable pathways and he showed some formal research that his team had done at ESCAP

(5) The policy implications, and what kind of policy options are available.

A few conclusions. Dr. Malik recommends us to focus on reducing setbacks; in other words, building resilience is as important as pursuing rapid GDP growth, or accelerating growth, because shocks leave long-term scars; and you lose your development gains over time. How do you build resilience? You invest in education, invest in health, and invest in social protection. He showed evidence why they are beneficial for the economies. Secondly, we need to rethink not just economic thinking but the actual development pathways. So, you should deliberately and simultaneously pursue resilience, inclusiveness and sustainability, not GDP growth first and then another consideration. The model that he proposed for the "build forward better" policy package on three areas to actually do that.

The full transcript of the presentation is available in Annex 1. The recording and presentation are available on the website at https://unstats.un.org/bigdata/events/2022/conference/agenda.cshtml

Panel discussion

This interactive panel discussion looked in more detail at how data can help with policy choices which governments need to make. It looked at how the gap between data producers and data users can be bridged. Dean Affandi from World Resources Institute Indonesia, shared on the importance of both policies and politics. He highlighted the importance of ensuring data is translated into a format that policy makers can both understand and use it easily. He also noted that the interdependencies of various policy spaces emphasize the need for quality and timely data to make effective and inclusive decisions.

Stefano Balbi, a researcher in sustainability science and natural resources management involved in the management and development of the ARIES (ARtificial Intelligence for Ecosystem Services) community highlighted how science was used in a timely way to tackle Covid and asked why we are not doing the same for other crises like climate change. He noted how data is too often compartmentalized and how we need to look for ways to integrate knowledge, bring data together, and increase access to data.

Hamza Malik, Director of Macroeconomic Policy and Financing for Development Division, UNESCAP, noted that the use of data as well as how data is presented is important. He highlighted the importance for policy makers to use data wisely in designing inclusive and ecologically sustainable policies.

The discussion looked at a range of issues from how governments can help countries to access private big data sources to provide collective goods through to how governments can have an important role in standardizing and harmonizing data and data rules. Having a strong narrative supported by data was also seen as being very important.

Food security

The second high-level panel was about food security and nutrition. Supporting the livelihoods of small-scale food producers, improving the resilience of food production systems and encouraging the sustainable use of natural resources are all key to fulfilling this mandate and achieving the Sustainable Development Goal in this area (SDG2), a world without hunger, food insecurity and malnutrition.

Dr. Carlo Cafiero, Senior Statistician and Economist, FAO

What Dr. Cafiero discussed at the conference was where we are in terms of informing policies that are intended to improve access to food for people. Before going into the details, he made a first premise, namely that "Food security" in a way has a different meaning than what is probably still commonly understood. By food security we essentially mean the outcomes at the individual level. So, when we discuss food security it is no longer talking just about food availability but rather about the individuals ability to access the food they need continuously and of the right quality.

The individual food security and the nutrition outcomes related to food is the result of a complex interaction of factors that happen at very different levels. Some factors that are more proximal to the people, some that are more distant, including phenomena that occur at the global level. So, this is the reason why we see food security as the result of many interactions. This is quite a departure from more traditional ways that see food security only as the adequacy of the total supply; so, you would just look at the food balance. It's more important to understand what happened with the food that is available and how different people can actually access it.

Another premise is that we define food security as composed of several dimensions and recently the definition has been expanded to include not only availability, access, utilization and stability, but also agency and sustainability to recognize the importance that people take control of what they eat of how they utilize the food.

The full transcript of the presentation is available in Annex 2. The recording and presentation are available on the website at https://unstats.un.org/bigdata/events/2022/conference/agenda.cshtml

Panel discussion

Prof. Martianto is a faculty member at the Department of Community Nutrition, Faculty of Human Ecology at IPB University Bogor. He indicated that his university works closely with the Indonesian Government's Food Security Agency and also with FAO to cover the complex Food system of the country. Together they developed what is called the Food Nutrition for Vulnerability Atlas. It is very important for Indonesia with our more than 17,000 islands to really understand what the problem is in

either food production, distribution logistic or the nutritional status. The food system needs to reflect that diversity.

Dr. Kadarmanto is the director of food crops, horticulture and estate crops statistics at BPS Indonesia. His team is preparing the 7th Agriculture Census of Indonesia, which will be carried out in May 2023. This census will provide information on the progress of the agriculture sector in Indonesia and on SDG indicators for agriculture. It will also provide social demographic information of the farmers. The Census will employ more than 350,000 enumerators and will enumerate more than 28 million farmers.

Dr. Cafiero added that nowadays food security is a very hot topic in all international discussion including the G20 and that from the perspective of data and information what the G20 also has pushed for is to really strengthen the coordination, to make sure that information, that is made available, is solid.

It was further pointed out that Susenas (the National Social Economic Survey of Indonesia) is very important to provide social and demographic indicators. The survey is conducted in March and September. At times it is important to also get information on the in-between months. New data sources, such as social media or crowdsourcing could help in estimating indicators for those missing months.

Another point which was made is that very high-frequency data, like daily indicators, may not necessarily give you a lot of information. It is important to distinguish between what is usual seasonality and variability of local production prices (even consumption habit) from what is truly an exceptional event. We may have more and more data (like daily data), but we need to be very careful on how we interpret those data. Welcome the possibility the technology provides, but it is very important to embed the data into models that can really link the pieces of information in a coherent way.

Access to relevant data

The third high-level panel was about the importance to have access to relevant data. In June 2022, the Conference of European Statisticians released a position statement on the access to data held by the private sector for purposes of official statistics. Among others, they called upon national governments and relevant ministries to support national statistical offices in their key coordinating role in the access to data held by the private sector for official statistical purposes, bearing in mind the critical role of high-quality, independent and impartial official statistics in evidence-based public debate and informed decision-making and taking into account the setup of the statistical system in a particular country.

Prof. Dr. Bertrand Loison, Federal Statistical Office of Switzerland

Prof. Loison presented about the work that was done in Switzerland regarding the creation of a data science competence center and about corresponding data strategy and data governance. He covered two topics: data analytics and data governance; first, he introduced the motivation why a data science competence center was created. Back in 2015, the message was quite clear: we are in the middle of a data revolution, and you need to anticipate the effect of this data revolution on your work, if you still want to be relevant in 20 years. In the UK they published a very interesting report asking the question: are we still measuring the right thing, if you want to understand our economy? We have technical challenges, and we have also data analytics challenges, and it was really the question of how can we, as official statisticians, react to this new world?

In Switzerland, a Data Stewardships Model was implemented in 2022. It is quite a complicated model, but it works. The Director General of the Swiss Statistical Office is also the Swiss Data Steward; so, he has a role at the table. He can indicate to which data his office would like to obtain access, and how these data can be accessed.

The full transcript of the presentation is available in Annex 2. The recording and presentation are available on the website at https://unstats.un.org/bigdata/events/2022/conference/agenda.cshtml

Panel discussion: access to mobile phone data

Mr. Matjaz Jug of Statistics Netherlands was the moderator of the final two panels of the day. The first panel discussed concrete examples of one specific data source, namely mobile phone data, where Indonesia is leading the way and demonstrates the potential of big data for official statistics.

The panelists were Ms. Shorena Tsiklauri of the national statistical office of Georgia, Mr. Alfian Manullang of Telkomsel, Indonesia, and Prof. Setia Pramana of Politeknik Statistika STIS and BPS Indonesia. As Prof. Pramana explained, implementing the use of mobile phone data (MPD) is not just a case of one unit in BPS; it requires collaboration with many other parts of BPS such as the IT department, which handle the technical aspects of the MPD, the department of Tourism statistics and the department of mobility statistics. Outside of BPS, they work together with the Ministry of Tourism and Creative Economy, and of course with the Mobile Network Operator, which is Telkomsel. The reason why BPS Indonesia was very interested in using MPD was because in certain remote areas of Indonesia immigration data or Border surveys do not work very well. Initially, MPD was used to estimate cross-border statistics (like foreign tourism) and later commuting statistics and domestic tourism were added.

Telkomsel is the biggest Mobile Network Operator in Indonesia with 170 million customers and almost 1 million cell towers. Telkomsel collects from its customer every movement and every call record (like SMS). Only the event is recorded, not the content. They gather also some demographic and social characteristics, while complying with regulations similar to GDPR. Telkomsel uses data internally to analyze product recommendation or customer engagement. In Indonesia, only about 50% of the people have access to the financial banking system. For the other 50% it is difficult to determine if they could receive a small loan, since they have no credit score. Based on internet data, Telkomsel created an alternative measure for people's financial behavior, which can be used as an alternative credit score to obtain small loans.

In the collaboration between BPS Indonesia and Telkomsel, Telkomsel initially provided some sample data to BPS Indonesia in a Sandbox. BPS then developed its algorithms in the Sandbox and subsequently Telkomsel ran the algorithms on the full dataset. BPS is taking a leading role in model development based on sample data and Telkomsel runs the model in its environment for all our customer and produces the aggregate output and also some quality assurance metrics.

The national statistical office of Georgia also has some experience using MPD. In 2016 -2017 a pilot project was implemented with the aim to develop and a complement existing methodology and statistics for the Information Society. One of the main players in this project was GNCC, which is the regulatory body for mobile network operators in Georgia. The project was implemented successfully with very nice results. To be successful it is essential to have solid coordination mechanisms between all stakeholders involved in the project. Further, data privacy is crucial, because MPD includes personal and

business sensitive information. Resources must be allocated both for human and technical capacity; and last but not least a successful project needs the support from high-level decision makers, who can make the project more sustainable; so political will is also important.

Panel discussion: partnerships with private sector

The second panel on access to relevant data covered partnerships with private sector. Mr. Matjaz Jug was again the moderator. Panelists were Dr. Veronika Vilgis of Dataprovider.com, Mr. Jack Fitzsimons of Oblivious.ai and Mr. Federico Sanson of UNHCR.

At Dataprovider.com, every single website on the internet gets indexed on a monthly basis and all the information collected from the websites is then structured into a database. Currently, over 200 data fields are available covering different categories such as location, technical or hosting information, and also marketing information and business name, business registry number etc. This done every month to see what happens on the internet over time. In 2015, Dataprovider.com started working with Statistics Netherlands for a study on the internet economy. Web data of Dataprovider.com was merged with the national business registry. In this way, we could come up with a classification of businesses based on their online activities. This partnership has continued ever since.

Partnerships with the public sector what works best if it's an open dialogue to learn from each other. Getting feedback is important for Dataprovider.com to validate the web data. No one actually knows how big the internet is. So, it's very hard to have a benchmark.

UNHCR collects large amounts of data that it uses for operational reasons, but these data can be reused for analysis. UNHCR would like the micro data to be used both internally and externally while making sure that the safety of the refugees is the top priority when sharing data knowing that they are already vulnerable. Making the data available to other organizations can deliver insights that can help the refugees. So, it's always a matter of finding the right balance between utility and risk. When we share the data currently, we mostly use statistical disclosure control methods through which we normalize the data, and we reach an acceptable level of risk. The Micro Data library of UNHCR contains currently more than 500 data sets.

The role of an NSO is to monitor the progress or regression of society across a range of different factors and also provide data to help inform government decisions. The more data you have, the more fine-grained it is, the more nuanced the information you can share, but at the same time you may start to overstep the privacy rights of your citizens. So, you have this balancing act between getting the utility of data and the privacy and governance constraints.

Privacy enhancing technologies is a field that started with statistical disclosure controls many years ago. More recently there's been a range of advances in two main areas. One we refer to as input privacy and the second we refer to as output privacy. Input privacy is all about how you can control (through encryption or secure enclaves) that the use of the data is restricted to processes that you trust between your organization and another. Output privacy is like the data disclosure controls in which you obfuscate or perhaps perturb the results of statistics, such that no one can reverse engineer the original inputs that you know may pertain to an individual for example or a business.

Day2: Dialogue with the statistical community

On the second day of the conference the UN Committee of Experts on Big Data and Data Science for Official Statistics (UNCEBD) held its annual plenary session to discuss its programme of work and to collect views of the members on the future direction. Subsequently, some of the UNCEBD task teams had organized several panel discussions regarding capacity development in the use of Big Data and data science for official statistics, regarding Big Data and the SDGs, use of mobile phone data, and regarding data science and the access to new data sources.

Recordings and presentations of the sessions on Day 2 can be found under the "Tuesday, 8 November" tab of https://unstats.un.org/bigdata/events/2022/conference/agenda

Plenary meeting of UNCEBD

The plenary meeting of UNCEBD was chaired by Mr. Ashwell Jenneker of Statistics South Africa. The agenda was as follows:

Item 1.   Overview of achievements and plans of the Task Teams
Item 2.   Overview of achievements and plans of the Regional and Sector Hubs
Item 3.   Overview of achievements and plans of the UN Global Platform
Item 4.   Structural and systematic collaboration between the Task Teams and the Regional and Sector Hubs
Item 5.   Strengthening collaboration between UNCEBD and Geospatial community
Item 6.   Terms of Reference of the Data Science Leaders Network
Item 7.   Data Access and the UN PET Lab
Item 8.   Yogyakarta declaration "the importance of relevant and timely data for urgent policy interventions" and report to the Statistical Commission

The first 3 items were for information only, items 4 to 8 were for discussion. Under the achievements of the task teams, we can mention the advances in the methodology with the UN PET Guide, the AIS Handbook and several guides on the use of mobile phone data, the case studies of the UN PET Lab and the projects on AIS data, the training courses coordinated by the task team on capacity development and the participation of the task teams at EXPO2020 and several other events in 2022.

Under agenda item 2, the Regional Hub for Big Data in China connected remotely and mentioned that they have the intention to make the hub into a global UN Center with expertise in the analysis of the Digit Economy and remote sensing for agriculture statistics. The hub further reported on its improved website, its capability building activities, its collaboration with UNCEBD Task Teams and the strengthening of its research, applications and communications around Big Data.

The hub in Rwanda had organized a few in-person workshops with several African NSO and some international agencies, as well as some virtual events on collaborative projects. The hub in Brazil has been very active with the organization of workshops, webinars and other activities, while establishing a good working relationship with the ECLAC Statistics Division. The hub in UAE had organized a conference during the EXPO2020 in Dubai in January 2022 and has been active in the International Data Science Mentoring programme (together with ONS, UK).

The update on the UN Global Platform consisted of an overview of the active projects and an overview of the maintenance of the platform in terms of services provided and of the financial situation.

Structural and systematic collaboration between the Task Teams and the Regional and Sector Hubs

As mentioned, the UNCEBD task teams prepared methodological guidelines, executed projects, developed training materials (including e-learning courses) and conducted training activities. The Regional Hubs are very well suited to host training activities. Together with the task teams, the hubs have organized workshops and webinars on the use of earth observations for agriculture statistics, the use of mobile phone data for a variety of statistics and indicators, the use scanner data and webscraping for prices statistics and the use of AIS data for maritime statistics.

Can the collaboration between the regional hubs and the task teams be improved? How can task teams better share their project outcomes, products & tools with Regional Hubs? How can Regional Hubs gain access to task team experts to support specific Big Data project work?

Several views were exchanged. It was summarized that the communication both within the UNCEBD as well as between UNCEBD and other organizations should be improved. The task team on Training, Competencies and Capacity Development will take on the leading role in improving this communication.

Strengthening collaboration between UNCEBD and Geospatial community

There is a proposal to strengthen the collaboration between UNCEBD and Geospatial community. This could be done by signing an exchange of letters between the Chairs of UNCEBD and UN Expert Group on Integration of Statistical and Geospatial Information (UN-ISGI), in which we express commitment to work on issues of mutual interest, like a quality framework for geo-referenced data points or joint training activities. This collaboration concerns the work of the task teams on earth observations, mobile phone data, AIS data and rural access to services.

There was strong support to commit to this collaboration. It was even expressed that if outputs of the mentioned task teams want to be meaningful, they should be integrated with available geospatial information.

Point of action is now that UNCEBD and UN-ISGI should get together and formulate in better detail how the collaboration will take place and which activities will be taken on first. Data science could be an integrator for this collaboration.

Terms of Reference of the Data Science Leaders Network

The Data Science Leaders Network (DSLN) convenes data science leaders of national statistical offices, including the leaders of UN Regional and Sector Hubs for Big Data. The draft Terms of Reference (TOR) of the DSLN provide the details on the purpose, the key functions and activities, the scope and responsibilities, the governance and reporting, the membership and the modalities of working of the DSLN. ONS, UK, presented the draft TOR.

The purpose of the DSNL is to enable strategic discussions on big data and data science issues, to share experiences and knowledge, to strengthen NSO decision making and leadership, and to accelerate research collaboration and technical partnerships. DSLN will provide input to the mandates of UNCEBD in the formulation of an overall vision, coordination, guidance, prioritization and direction in the area of data science for official statistics. The expected outputs are concrete strategies and roadmaps with a view to achieve coherent and integrated programmes of work in this area. The activities of DSLN will be Sprints on priority topics, which will allow fast results and allow addressing emerging issues.

Regarding the governance, it was proposed to have a dedicated Bureau, which will initially be chaired by the UK. Proposals for Bureau members were UK, Switzerland, Canada, China, Denmark, Netherlands, Rwanda, Brazil, Indonesia, UAE, IMF, UNECE, UNECA and UNSD. Further, several topics were mentioned for the Sprints, such as "Understanding the position of Data Science within the wider Government set up", "How to support NSOs to evolve to the next level of data science maturity", or "Requirements for and short- and longer-term benefits of setting up a data science team in an NSO".

During the discussion, the importance of capacity building in data science, especially in developing countries, was underscored. Some institutes, like SCAD in Abu Dhabi, UAE, mentioned that they offer e-learning courses in data science, which can be used by other institutes. Members were invited to send comments on the draft and the possible topics to UNSD.

Data Access and the UN PET Lab

Obtaining the right source data has always been a priority for official statistics. Increasingly, we depend on secondary sources (both from governments and private sector) and Privacy-Enhancing Technologies (PETs) can make a difference in obtaining such sources. With the UN PET Lab we can (and should) now make the step from theoretical experiments to practical applications. We have to identify sources and use cases to create synergy between the task team on Facilitation of Access to Privately held Data and the UN PET Lab, to make real progress.

Several questions were put in front of the UNCEBD members, namely for which high-value data sources could PET help to unlock the data for statistics production? Which use cases could be implemented with existing partners to show a PET-based approach works? Are there any other barriers (for example non-technical) that we should specifically consider?

The discussion regarding using real-life data sets for testing of PETs focused on balancing the relevance of data (for example from health sector) with the feasibility of being allowed to use the data (given legal constraints). The task team on PETs will both test PETs from a technology point of view as well as investigate the legal basis of being allowed to use sensitive data when PETs are used. The UN PET Lab will also offer services to statistical institutes, if they are interested to start using PETs.

Yogyakarta declaration and report to the Statistical Commission

The structure of the Yogyakarta declaration would include the mandate of UNCEBD, the Cape Town Global Action Plan for SDGs, the Bogota, Kigali and Seoul declarations, and a reference to the theme of the Conference, to the Regional and Sector Hubs, the collaboration statistical and geospatial community, encouragement of participation in DSLN, and a call for using PETs in official statistics. In general, the members supported the declaration, while stressing that the points of the declaration should be focused.

The recording of the plenary meeting starts here: https://youtu.be/_98ngGFBtJI?t=173

Capacity development and Regional Hubs

This panel discussed capacity development for the use of Big Data and the role that Regional Hubs play in this process. The regional and sector hubs provided an overview of the work they have been doing so far and upcoming plans. The understanding of needs by NSOs and other entities exploring the use of Big

Data is a key concern for the regional hubs, as these needs should form the basis for work programmes to be established and for the identification of possible partners to support the operations. The representatives of the regional and sector hubs elaborated on the methods and tools they have used to obtain the proper understanding of those needs.

The actual methods employed by the hub to facilitate capacity building and knowledge exchange were described by the hubs and showed interesting differences based on the state of development of the hubs at this point. The tools, such as webinars and focused training sessions, reflect also the needs identified in the gathering stage and will be further evaluated to enable the hub to reach the widest possible audience in these efforts.

The discussion with the audience showed great appreciation by countries of the efforts undertaken by the hubs. It also highlighted the strong need to better publicize the activities undertaken to allow NSOs and other country agencies to be aware of products already produced and being available, as well as to make use of capacity development activities offered and planned.

The recording of the session on Capacity Development and Regional Hubs starts here: https://youtu.be/_98ngGFBtJI?t=18267

Big Data and SDGs

The panel on Big Data and the SDGs presented status and possibilities of the application of non-traditional data sources both for monitoring the SDGs, but also in a wider context of producing statistics as seen from different perspectives - national and international. The overarching goal was to provide an overview of the ongoing activities in this regard, a concrete input to further work on Big Data and statistics, and some recommendations on the concrete steps to move work on Big Data and statistics forward.

The presentations underlined to potential of Big Data for production of statistics but also identified challenges and learning points. The following challenges were mentioned: (a) understanding of the problem – what is it to be monitored, i.e. the link between the data and the indicator; (b) understanding the data; (c) getting the right skills; (d) compliance of data with international standards; (e) data privacy issues.

Furthermore, the attention was drawn to the differences between producing statistics in a 'traditional' way and with the use of Big Data. The latter is calling for more work regarding the structure of the data, the need for a diversified capacity building, and the need for complementing Big Data with other data sources. It was further emphasized that an ongoing analysis of what has worked and what has not worked is particularly important when working with Big Data. The current learning points show that systems for compiling statistics using Big Data should be built together with users, built for scale in a gradual approach, and be subject to a thorough and consequent planning.

The recording of the session on Big Data and SDGs starts here: https://youtu.be/_98ngGFBtJI?t=21445

Mobile phone data for official statistics

This interactive panel on the use of mobile phone data for official statistics highlighted the importance of cooperation in implementing mobile phone data (MPD) projects. The session, chaired by Esperanza Magpantay (ITU) and composed of representatives from national statistics office (Statistics Georgia, BPS

Indonesia), Ministry of ICT (DICT Philippines), Flowminder and Telkomsel Indonesia, identified the challenges faced by NSOs related to access to data, lack of capacity and human resources, and challenges with infrastructure to implement use cases in different areas of statistics. The work of the UN-CEBD Task Team on Mobile Phone data (MPD) could help address the issues highlighted, including the use of international agreed methodologies, and working closely with national stakeholders, particularly the mobile operators in the country. Five Methodological Guides on the use of Mobile Phone Data (2022) were launched during the session, covering areas such as tourism, migration, dynamic population mapping, information society, and displacement and disaster statistics. Countries and practitioners are encouraged to use the Guides when implementing MPD use cases.

The recording of the session on Mobile phone data for official statistics starts here: https://youtu.be/_98ngGFBtJI?t=25378

Data Access and Data Science

The panel on Data Access and Data Science focused on the central question 'If we obtain data from 'outside', how much information/influence/control do we need to make sure that we understand its quality and can explain how we compile our statistics?' The question was introduced by reminding the audience of the role of quality in the UN Fundamental Principles of Official Statistics and the many aspects of quality that need to be taken into account. The panelists were Mr. Daniel Boller and Mr. Joseph Bulan of the Asian Development Bank, Mr. Eric Deeben of the Office for National Statistics, UK, Ms. Evest Defiana of Positium, Ms. Muna Al Mazrouei of the Federal Competitiveness and Statistics Centre of the UAE and Ms. Veronika Vilgis of Dataprovider.com with Mr. Barteld Braaksma of Statistics Netherlands as the Chair. The panel thus consisted of a diverse mix of people from different regions and different types of organizations, including the private sector.

Each panelist introduced the topic from their own perspective through a short pitch, describing a concrete example. Examples covered sources like satellite data, online (website and social media) data and mobile phone data. The panelists highlighted several aspects like data processing, data integration, visualization and the societal challenges (such as poverty) concerned. The set of statements already provided a rich portfolio to start a discussion where the audience was actively invited to contribute through interactive Mentimeter questions. The questions related to quality considerations in public-private partnerships, key challenges for quality assurance in such settings, capacity building issues and the changing role of official statistics.

Important insights included that ensuring (processing) continuity and creating transparent processes are of the utmost importance when dealing with privately held data and data science methods; traditional quality dimensions like accuracy, relevance and timeliness remain important (although the balance between them may shift); and that we should aim for a multi-source approach where the strengths of one source may compensate the weaknesses of another. A specific observation was that the private sector representatives in the panel showed a good understanding of the quality requirements of official statistics, which is a good basis for building trust and creating productive public-private partnerships.

The recording of the session on Data Access and Data Science starts here: https://youtu.be/_98ngGFBtJI?t=28742

Workshops

On Days 3 to 5 (Wednesday, 9 November 2023, to Friday, 11 November 2023) 10 3-hour training sessions were organized in parallel in two conference rooms. The schedule is given in the table below. The purpose of the workshops was to introduce a variety of topics of use of Big Data and data science to the conference participants and to provide some hands-on experience as well. The topics were privacy-enhancing technologies, use of AIS data for maritime transport statistics, use of Earth Observation data for agriculture statistics, use of mobile phone data for official statistics, and the use of Machine Learning algorithms for production of official statistics, including getting access to non-traditional data sources.

The recordings, presentations and additional materials can be found at
https://unstats.un.org/bigdata/events/2022/conference/workshops

Day	Room A	Room B
Nov 9 09:00 – 12:00	UN PET Lab Overview of the use of several PET techniques, which could be use during the PET Lab Hackathon Competition	AIS Data (part 1) − Introduction to the AIS Task Team and the work programme (presentation) − What is AIS and using UNGP AIS (hands-on) − Optimizing AIS data extraction (hands-on)
Nov 9 13:30 – 16:30	EO Data (part 1) − Standardization of EO Methods; ML approaches for Land cover mapping given in-situ data scarcity − Sharing of country experiences	AIS Data (part 2) − New CSO Ireland frontier indicators based on AIS data (presentation and demo) − Identification of ports/berth using ML (presentation and demo) − Discussion on potential use cases using AIS data (round table)
Nov 10 09:00 – 12:00	EO Data (part 2) − 50 X 2030 initiative and in-situ data requirements. Best practices in geo-referencing − Building geospatial data science training courses based on user profiles − Geospatial data sharing using UNGP: present solutions and future directions	Data Access and ML (Part 1) − Examples of the use of online (website) data for statistical purposes such as the impact of the digital economy and e-commerce, analysis of online job advertisements, online signals of COVID effects. − Exploration of the database and search engine of dataprovider.com, consisting of half a billion websites in more than fifty countries – hands-on
Nov 10 13:30 – 16:30	MPD Data (part 1) Introduction and Launch of the Guidelines on use of MPD data for: Dynamic Population and census; Tourism; Migration; Information Society; Displacement & Disaster; Transport	ML (part 2) Overview of use of Machine Learning in NSOs

Nov 11 09:00 – 11:45	MPD Data (part 2) How to implement an MPD pilot in 9 steps, Collaboration with Regional Hubs Break-out groups on use of MPD in developing countries	ML (part 3) Hands-on examples of use cases with Machine Learning in NSOs

UN PET Lab

The UN PET Lab workshop brought together four experienced speakers to educate the audience on privacy enhancing technologies (PETs) and their impact on official statistics. The audience was composed primarily of data scientists with a background in official statistics, making for a knowledgeable and engaged crowd. The session received a full attendance of approximately 80 in-person participants and was also made available online.

The workshop opened with a few words from Prof Setia Pramana of BPS Indonesia who emphasized the growing importance of PETs in official statistics and the recent work done in Indonesia to bring together telecom providers for better tourism statistics. This set the stage for the workshop, which was designed to introduce the audience to PETs and the work of the UN PET Lab.

Jack Fitzsimons, the CTO of Oblivious.ai, started the workshop by introducing the audience to the concepts of input and output privacy, such as homomorphic encryption, secure multiparty computation, and differential privacy. He linked these concepts to the open challenges faced by government organizations when collecting, processing, and disseminating sensitive information.

Matjaz Jug, Head of the CIO office at Statistics Netherlands, then spoke about the origins of the UN PET Lab and shared a number of PETs case studies, motivating the audience to learn more about these technologies. Federico Sanson, from the Data Disclosure Team at the UNHCR, spoke about the need for data in their organization and their goals to share openly and responsibly with internal and external stakeholders. This is a core part of their Data Transformation Strategy 2023.One major action of this strategy was to build the microdata library, which is publicly accessible, but which requires users to request access to individual datasets for manual review. They currently host over 500 datasets and this number is growing.

Finally, Pauline Maury-Laribiere, a data scientist from Switzerland, introduced the dataset, which was used in the UN PET Lab Hackathon competition, which focused on the impact of COVID-19 on refugees. She detailed how she had endeavored to make sure the dataset would be engaging for the competition itself, while ensuring the privacy of individual records pertaining to individual refugees were kept private and anonymous. Jack Fitzsimons concluded the workshop by giving a brief overview of the Python interface that participants could use to join and compete in the competition.

The audience was engaged throughout the workshop and several participants asked questions ranging from the technical maturity of different PET solutions, the technical aspects of them, through to how the technologies related to legal and regulatory aspects of data privacy. There were also questions as to whether or not these technologies were suitable for smaller datasets and their connection to the data disclosure controls already in place by most NSOs and UN affiliated organisations.

Overall, the workshop was an educational and engaging experience for all who attended. The speakers provided insightful information on the topic and encouraged the audience to continue learning and exploring the world of privacy enhancing technologies.

AIS Data

The workshop on the use of AIS data aimed to raise awareness of various use cases of AIS data, including transport and emissions indicators, and to undertake hands-on exercises in extracting the data from the UN Global Platform. It was attended by around 75 people and facilitated by the members of the UNCEBD task team on AIS data. It was started by introducing the Task Team and followed by explaining the concepts of AIS data, including their limitations. Then, the facilitators described a procedure for requesting access to the UN Global Platform and demonstrated a step-by-step walkthrough on AIS data extraction. All participants were given temporary access codes; therefore, they could follow the walk-through on their own laptops and obtain the data using PySpark.

The second session focused on the experimental maritime indicators developed by the Central Statistics Office of Ireland, such as port calls. The speaker described in detail the justification for using the AIS data for maritime indicators – notably its reliability and accessibility, data cleaning and processing methodology, and the way forward. Another topic in the second session is using machine-learning techniques to identify ports and berths boundaries – based on the actual AIS messages transmitted by the ships. The same method can also be utilized to identify ships' waiting areas or fishing zone that may not be visible otherwise (using a map). Finally, the participants greatly appreciated the Task Team's work on AIS data and the UN Global Platform, notably for providing the official statistics community with open access to AIS data and its computing platform.

EO Data

The workshop on the use of earth observations (EO) for agriculture statistics was delivered over two sessions with presentations by Mr. Lorenzo De Simone, special advisor to the Chief Statistician of the UN Food and Agriculture organization, Mr. Talip Kilic, Senior Economist at the Development Data Group of the World Bank, and Prof. Kerrie Mengersen of the Queensland University of Technology, Brisbane, Australia. They work together in the UNCEBD task team on the use of EO for official statistics, especially for agriculture statistics. More information on the task team can be found at
https://unstats.un.org/bigdata/task-teams/earth-observation

In the first session Mr. Lorenzo De Simone presented about standardization of methods to produce official land cover and crop statistics using EO data. He presented a variety of existing applications such as the crop and landcover mappers, the global agro-ecological zoning portal, the agricultural stress index or the water productivity portal. He then showed in detail the heterogeneity in landcover land-use applications due to differences in classifications, geo-referencing and processing of the images. He made the case that a quality assurance framework for EO data, including standards and governance mechanism, should be identified. As an example, he showed the methodological and standardization work done on SDG indicator 15.4.2 (Mountain Green Cover Index). In a second presentation Lorenzo addressed sustainable methods for national land cover and land use mapping in the context of scarcity of in-situ data and project resources. Crop type maps are a "specialization" of the land cover class "agriculture". In-situ data are essential for the automatic classification of satellite images into land cover classes. Some "fallback" solutions in case of scarcity of in-situ data are reuse of older or similar data, or

use of data frugal classification algorithms. The presentation was a followed by a Q&A discussion with the workshop participants.

In the second session, Mr. Talib Kilic presented on the "Integration of Surveys and Satellites for Crop Type Mapping and Crop Yield Estimation". The objective of this 50X2030 program is to conduct rigorous research to answer high-priority questions in microdata collection, analysis and dissemination to support scaling up accurate integrated satellite-survey applications, while focusing on an attainable set of agricultural outcomes – production, cultivated area, and yield – for **a minimum of four cereal crops** - namely maize, rice, sorghum, and wheat. The main output of this program are *Guidelines* for georeferenced microdata collection and processing for calibrating and validating remote sensing models for high-resolution crop type mapping and crop yield estimation. Additional outputs are working papers and peer-reviewed journal articles, high-resolution seasonal crop area and crop yield maps for select crops and countries covered by research, and geo-referenced training datasets for downstream use in EO applications. He then showed in detail some main achievements, like the Maize area mapping projects in Malawi and Ethiopia.

In the second session, Prof. Kerrie Mengersen also made a presentation. She gave an overview of the development and use of a personalized training program. Starting from different user profiles and knowledge areas (like remote sensing, GIS, data management, data analytics or data visualization) you could find specific online learning courses for a personalized training program in, for example, remote sensing for agriculture statistics. These courses could further be available at different levels of expertise (from beginner to advanced). A first personalize self-learning system has been made available. The next steps include an extended audit of trusted Online Training courses, refining the prototype Personalised Self Learning System and develop some certification.

Data Access and Machine Learning

The Data Access and Machine Learning (ML) workshop was delivered over three sessions to 70+ participants by a team of experts from the UNSD, UNECA, UNESCAP, UK ONS, Statistics Indonesia, Statistics Netherlands, Dataprovider.com, the Regional Hubs for Big Data in China, Rwanda and UAE, and representatives from the private sector with expertise in data sharing.

The workshop's objectives were:

- Understand the opportunities of online data and gain hands-on experience with using a (big) database of websites
- Learn about the different applications of machine learning for big data
- Understand how machine learning is being used in official statistics
- Identify the key factors that need to be in place in organisations in order to begin implementing a machine learning project
- Develop ideas for collaborative ML research projects for each UN regional hub.

Session I was led by Barteld Braaksma (Statistics Netherlands) and Veronika Vilgis (Dataprovider.com) and focused on the statistical use of web data. In the first (theoretical) part, examples of the use of website data for statistical purposes were discussed, including measuring the impact of the digital economy and e-commerce and an analysis of online job advertisements. In addition, use cases concerning the potential of social media were described. Next, some applications of the database of Dataprovider.com for official statistics were highlighted, while also allowing a sneak peek of applications

beyond statistics, such as an analysis of online signals of COVID effects carried out together with ETH Zurich and VU Amsterdam. The second (practical) part showcased the potential of the huge database and search engine of Dataprovider.com. The database consists of half a billion websites in more than fifty countries. For each website, over two hundred attributes are systematically derived and stored. All workshop participants received guest accounts and were guided through the data platform with a carefully crafted series of assignments.

Session II followed with presentations from the UK ONS and Statistics Indonesia on ML for big data in official statistics. Two use cases were then discussed in further detail across all participants addressing poverty through geospatial big data (Statistics Indonesia) and economic modelling (UAE team).

Session III started with presentations from the Regional Hub for Big Data in China. In order to focus discussions about ML innovations in official statistics on specific regional issues, participants were split into 4 subgroups according to their participating regions. The session ended with feedback from each subgroup lead and agreement of next steps across all participants. The topic/questions presented to the participants included:

- What is the business need for ML in your region?
- What improvements does ML promise for quality, efficiency, relevance, timeliness of output?
- What stage of readiness does your organisation need to be at for implementing ML projects (organisation experience, data access, tools and training, infrastructure)?
- What are the key challenges?
- What research topics are of common interest to statistical organisations in your region?
- What kind of machine learning project is most feasible for a joint project?
- What types of data sets could be used?
- What kind of training would be needed?
- What other resources and support would be required?
- What sort of outcomes would be expected?

This led to a successful and inclusive workshop, where participants had a unique opportunity to share solutions and concerns to common challenges in ML for official statistics in their corresponding regions, exchanging their knowledge and experiences and agreeing a clear plan and ways forward. The feedback from delegates was excellent, with a range of proposals for more regular activities in this format to strengthen connections between UN Regional Hubs, share resources to tackle common challenges and accelerate innovation in official statistics globally. A summary of discussions, outcomes and next steps for each region as a result of the ML workshop is presented below.

Asia & Pacific Region (led by UNESCAP)

- Bangladesh and Philippines were represented, and both seemed keen to undertake a collaborative satellite imagery and agriculture project with support offered by UNESCAP. Contact was made with both NSOs and further support will be required.
- UNESCAP is keen to progress the concept of a collaborative ML project with support of Statistics Indonesia. UNESCAP is interested in taking a step back, engage with a wider group of NSOs from across the region, identify their needs and gaps, and work up a collaborative project plan. Experiences will also be shared with the Regional Hub in Africa, who launched a collaborative work with NSOs in the region on web-scraping prices for CPI project.

Africa Region (led by UNSD)

- Delegates from NSOs of Senegal, Sierra Leone and Zimbabwe were present and offered to explore a project for automatic classification of occupation text. Senegal is keen to also explore and build skills for web scraping prices for CPI. The Hub was already scoping out a collaborative project for this. A Senegal team (team of 5) has now been brought into that project. Working alongside Ghana, Namibia and Tunisia, the team will learn from each other while working on their own in-country websites/online retail outlets. Instructors from the UK ONS, UN, ESSnet and ML international network have been secured to support this capacity building project.
- The Africa Hub is looking for some resource to support the creation of a new collaborative NSO project for automatic classification of occupation text to be run alongside the web-scraping prices collaborative project. In the absence of any extra resources, this may have a delayed start, potentially June 2023 onwards.

MENA (led by the UAE Regional Hub)

- Participants from Statistics Palestine, Statistics Abu Dhabi and Statistics Qatar were present and offered to explore a project for collecting tourism data and understand tourism spend.
- The team was keen to start a mobile phone data (MPD) project and is now building plans for this, which could be internal to UAE or a collaborative NSO project managed through the Regional Hub in the UAE. Further discussions are in progress.

Mobile Phone Data

The UNCEBD Mobile Phone Data (MPD) task team was established in 2015. The objectives of this task team are to develop methodological guidance, prepare training materials and organize activities for outreach and capacity development. The MPD Task Team organized a workshop during the 7th International Conference on Big Data that took place in from 7 to 11 November 2022 in Yogyakarta, Indonesia.[7]

The workshop was attended by around 70 participants, including representatives of national statistical offices, government, academia and NGOs from all regions. Participants from other UN agencies and regional organizations also attended the meeting. More than 60 % of the participants were women. The Brazil, Rwanda and UAE regional hubs were also present in this workshop.

The objective of the workshop was to present the five *Methodological Guides on the use of Mobile Phone Data (2022)*, on the use of mobile phone data in different areas of statistics, such as tourism statistics, migration statistics, displacement and disaster statistics, dynamic population statistics, ICT statistics, and transportation statistics.[8] The Guides elaborate on the different applications, data

---

[7] https://unstats.un.org/bigdata/events/2022/conference/workshops/index.cshtml.
[8] https://unstats.un.org/bigdata/task-teams/mobile-phone/index.cshtml. The Methodological Guide on Transport statistics is expected to be launched in July 2023.

sources, access to mobile phone data and partnership models, methods used in data processing and the methods to ensure quality of the calculated indicators, and country use cases. The Guides were officially launched during the interactive high panel organized by the MPD Task Team held on the second day of the conference.[9]

The details of the Guides were presented by the sub-group leads. The presentations highlighted the challenges in traditional sources of data which include timeliness, costs, burden to respondents, among others and which MPD could address. The presentations emphasized that MPD could offer opportunities by supplementing and complementing information for census processes, including in census preparation and execution, population projections, measuring possible undercounted populations, and providing disaggregated data at regional or provincial levels. At the same time, the pilots that were done on data collected using household surveys such as on Internet use, migration, and tourism confirmed that MPD could be used to complement or supplement traditional sources. For the case of displacement and disaster statistics, the presentation highlighted that aside from statistical production, MPD analysis could be used to facilitate timely response when disaster strikes to enhance preparedness and response of the authorities and those concerned. Country use cases were also presented to highlight the feasibility of replicating the experiences in other countries.

Based on the different Guides, the MPD task team worked on the awareness raising course. The course, which was presented during the workshop, aims to introduce the key concepts underpinning the use of MPD, highlight the opportunities presented by MPD, explain the strengths and limitations of MPD, and outline how to run a successful MPD project at policy makers and general public.[10] The course will be available online at the MPD Task Team website and will be delivered through the regional hubs.

During the discussions that followed, participants appreciated the launch of the Guides and emphasized the importance of learning from experiences of others. Challenges related to lack of capacity in NSOs, and data access were raised by majority of the countries present. It was suggested that synthetic data would be useful so that countries could use those data to learn the methods outlined in the Guides. The MPD task team agreed to explore the feasibility of providing synthetic data as part of the training materials to be provided to countries, in collaboration with the regional hubs. Data access mechanism, such as the one implemented in Indonesia, could be used as a model for countries who are planning an MPD project.

---

[9] https://www.youtube.com/watch?v=_98ngGFBtJI&t=25378s.
[10] The course will be available at the Moodle platform, see link XX.

2022 UN Big Data Hackathon

The UN Committee of Experts on Big Data and Data Science for Official Statistics (UNCEBD), the UN Major Group for Children and Youth (MGCY), and Statistics Indonesia joined forces in hosting the 2022 UN Big Data Hackathon with the challenge to develop ideas and solutions addressing global issues and helping achieve the Sustainable Development Goals (SDGs). Over 1500 data scientists and statisticians from more than 70 countries competed in the 2022 UN Big Data Hackathon from November 8th to 11th of 2022. Out of the 350 teams that competed, 165 teams successfully completed and submitted a project that addresses a specific issue in relation to the theme. In addition to the 2022 UN Big Data Hackathon, we also organized the UN PET Lab Hackathon competition which attracted an additional 196 teams, representing 328 individuals from 35 countries.

The UN Big Data Hackathon and the PET Lab Hackathon were both opened during the same ceremony on 8 November 2023. This ceremony can be viewed here: https://youtu.be/_98ngGFBtJI?t=16342 where the official start at 14:00 in Yogyakarta, Indonesia is here: https://youtu.be/_98ngGFBtJI?t=17986

The hackathon provided a platform for young people to use a wide variety of data to help solve global challenges. Rising energy prices are accelerating the cost-of-living crisis and sustaining the vicious cycle of constrained household budgets, food insecurity, energy poverty, and growing social unrest. More than ever, efforts to achieve the 2030 Agenda for Sustainable Development should be intensified. The event can serve as a channel for data science experts to communicate and serve as a venue for youth worldwide to learn together.

A total of 350 teams participated consisting of 49 big data track teams and 301 youth teams. Based on their preferences, teams competed in three on-site locations: Indonesia, UAE, and Brazil, and remotely for 72 hours. They were given objectives to help solve global challenges using a wide variety of data. They were also provided access to the UN Global Platform, AWS Event Engine, and ESRI. In the end, the teams must submit the following: presentation, video, and code.

The hackathon was an overall very positive experience for most participants. The newly added series of webinars (four in total) was a success since 89% of the hackathon participants enjoyed them. In addition, some feedback was received with regards to improving some aspects of the logistics as well as the access to datasets. Finally, 73% of the participants are most likely to participate in another edition of the hackathon.

More information can be found on the website at:
https://unstats.un.org/bigdata/events/2022/hackathon/

Youth Track

301 teams from all around the world competed in the Youth Track, of which 123 teams submitted their final project solutions. The first 3 winning teams in the competition were the following.

1st Overall Outstanding team : Data Eye (Lebanon) : **Spoonful**

The team introduces the platform "Spoonful" which aims to optimize food distribution programs in Lebanon by identifying population vulnerability by ensuring the right people are reached and optimizing food parcel price without compromising quality and nutritional targets. Identification of vulnerable populations is done by training machine learning models using night time light satellite data to assess poverty and vulnerability as has been supported by literature in recent studies. The application tries to maximize the number of households that can be served given a set of resources by optimizing the food box price by mapping commodity brands of more affordable brands without compromising on quality. This is achieved by web-scraping of prices performed on supermarket websites in Lebanon.

Using the Spoonfuls recommendation the web app provides the optimized subsidy basket price and optimal price of household served given the budget allocated. The application aims at increasing food subsidy budget lifetime value and help to reach more people who are truly in need in Lebanon and potentially all over the globe.
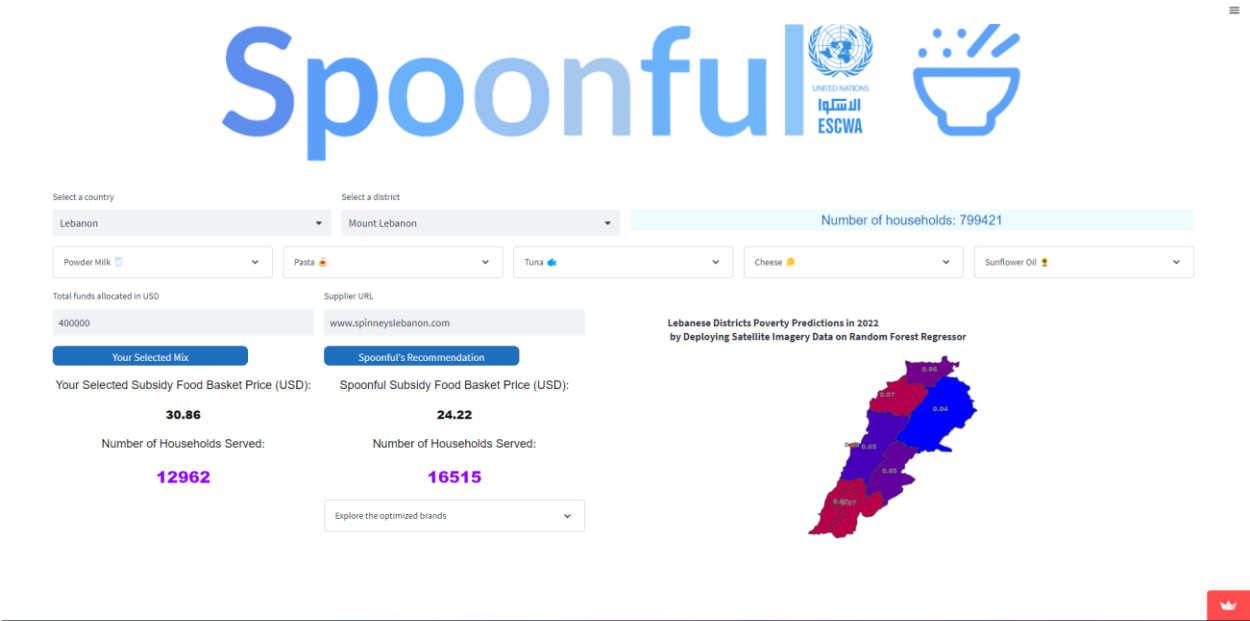
2nd Overall Outstanding team : M3GA (Philippines) : **Crop Diversification**

Aiming to reduce the impact of food supply shocks and inflation in food prices in Philippines due to the increased instances of natural calamities, team M3GA uses a network science approach to help policy makers to manage crop supply using crop diversification. This is achieved by identifying areas similar to each other but are geographically distant so natural calamities do not cripple supply of essential goods. Using cop volume production data and food prices data of the Philippines the team creates a crop profile for every spatio-temporal pair and nodes connected to their crop profile similarity. A condition of nodes being kept 250 km apart, which is the diameter of a typhoon, is added and used to identify potential planting alternatives. Using the network the team identifies connected nodes to check region quarter crop profile pairs that are alternatives to each other and optimizing this network can help in avoiding supply shocks and surging prices of food supplies incase of a natural calamity.

The resulting network shows pairs with the same spatial location but different quarters.

Philippine network is as such because there are no volatile seasonal changes. European countries may have greater diversity.

node1	node2	weight
REGION XI (DAVAO REGION)_Q3	REGION XI (DAVAO REGION)_Q4	60
REGION VI (WESTERN VISAYAS)_Q1	REGION VI (WESTERN VISAYAS)_Q2	54
REGION XI (DAVAO REGION)_Q2	REGION XI (DAVAO REGION)_Q4	52
REGION XII (SOCCSKSARGEN)_Q2	REGION XII (SOCCSKSARGEN)_Q4	50
REGION XI (DAVAO REGION)_Q2	REGION XI (DAVAO REGION)_Q3	50
REGION VI (WESTERN VISAYAS)_Q1	REGION VI (WESTERN VISAYAS)_Q4	49
REGION XII (SOCCSKSARGEN)_Q1	REGION XII (SOCCSKSARGEN)_Q4	48
REGION XII (SOCCSKSARGEN)_Q1	REGION XII (SOCCSKSARGEN)_Q2	47
REGION X (NORTHERN MINDANAO)_Q3	REGION X (NORTHERN MINDANAO)_Q4	47
REGION I (ILOCOS REGION)_Q1	REGION I (ILOCOS REGION)_Q2	46

3rd Overall Outstanding Team: Team 3994: **ArgAI**

Using AI driven ArgAI technology, Team 3994 tries to increase agricultural productivity to increase food supply, reduce agricultural inflation, reduce overall poverty and malnutrition. The team's ArgAI platform
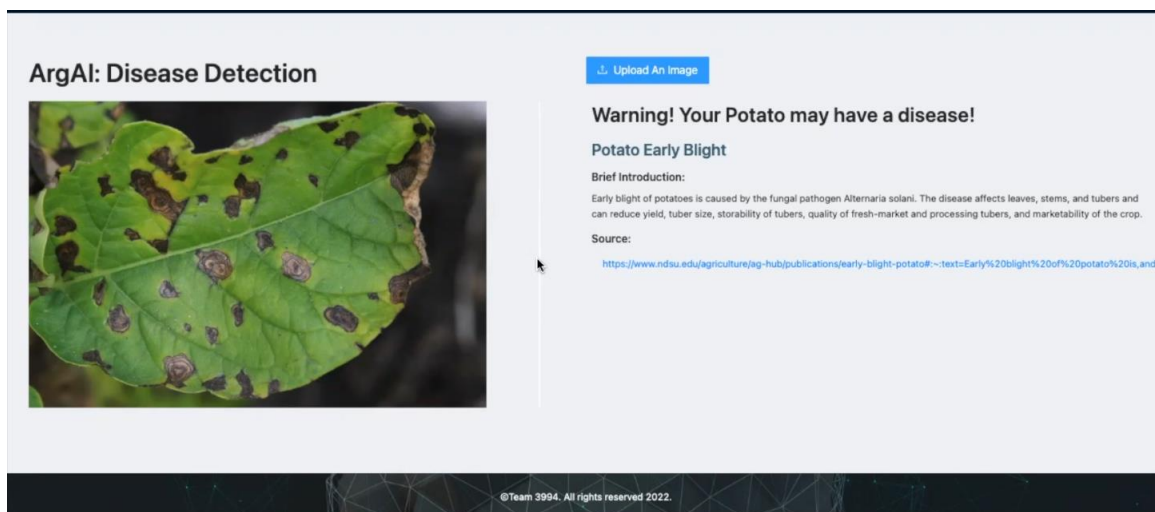
uses different data sources such as satellite imagery, cropland data, weather data etc, to train their AI system which provides a yield forecast and data based decision. ArgAI takes a soil image and using AI analysis provides a crop recommendation for the respective soil. ArgAI also uses a similar technique to detect crop diseases and provide recommendations based on the crop image.

Big Data Expert track

49 teams from all around the world competed in the Big Data Expert Track, of which 37 teams submitted their final project solutions. The first 3 winning teams in the competition were the following.

1st Overall Outstanding Team: Cerdasdata (Indonesia): **Monitoring Shipping Networks**

Using the AIS, port data, and official statistics datasets, the team aims to develop a system to monitor global shipping networks and to conduct analysis on the relationship between global shipping networks and statistical indicators of the world economy. Focusing on the current conflict, the team picked the Baltic and Black Seas as the target zone to analyze the impact of the Ukraine war. One of the conclusions was that the war does impact not only the warring nations but also the other countries within the shipping networks in the region. The table below shows the port ranking of selected countries

Area	Country	Median Rank 2021	Median Rank 2022	Diff
Baltic	Sweden	5	6	-1
Baltic	Germany	6	8	-2
Black	Turki	12	16	-4
Black	Greece	27	28	-1
Black	Ukraine	30	57	-27
Baltic	Finland	38	41	-3
Black	Romania	48	35	13
Baltic	Norway	52	68	-16
Other	Cyprus	67	92	-25
Baltic	Russia	73	162	-89
Other	Italy	87	180	-93
Other	Montenegro	88	86	2
Baltic	Denmark	92	103	-11
Baltic	Estonia	98	83	15
Other	Egypt	123	101	23
Other	Lebanon	130	110	20
Other	Israel	131	149	-18
Black	Georgia	156	145	11
Baltic	Latvia	167	115	52
Black	Bulgaria	184	202	-18

2nd Overall Outstanding Team: IMF-Big Data (IMF): **Saving Whales**

Due to increased ocean traffic encroaching on whale habitats, the team looked at the issue of great whale mortality due to collision with ships. Broadly, they aim to support international climate action by measuring the social cost of deadly vessel strikes on great whales, an important carbon sink. And to use the outcome to inform the design of whale conservation policies. The team developed a whale mortality

model based on several criteria, such as the size and speed of the ship and the whale, the ship's transit time, number of whales in the area. In addition to AIS and ship register data, they utilized the whale density dataset for US east coast and southeast Canada
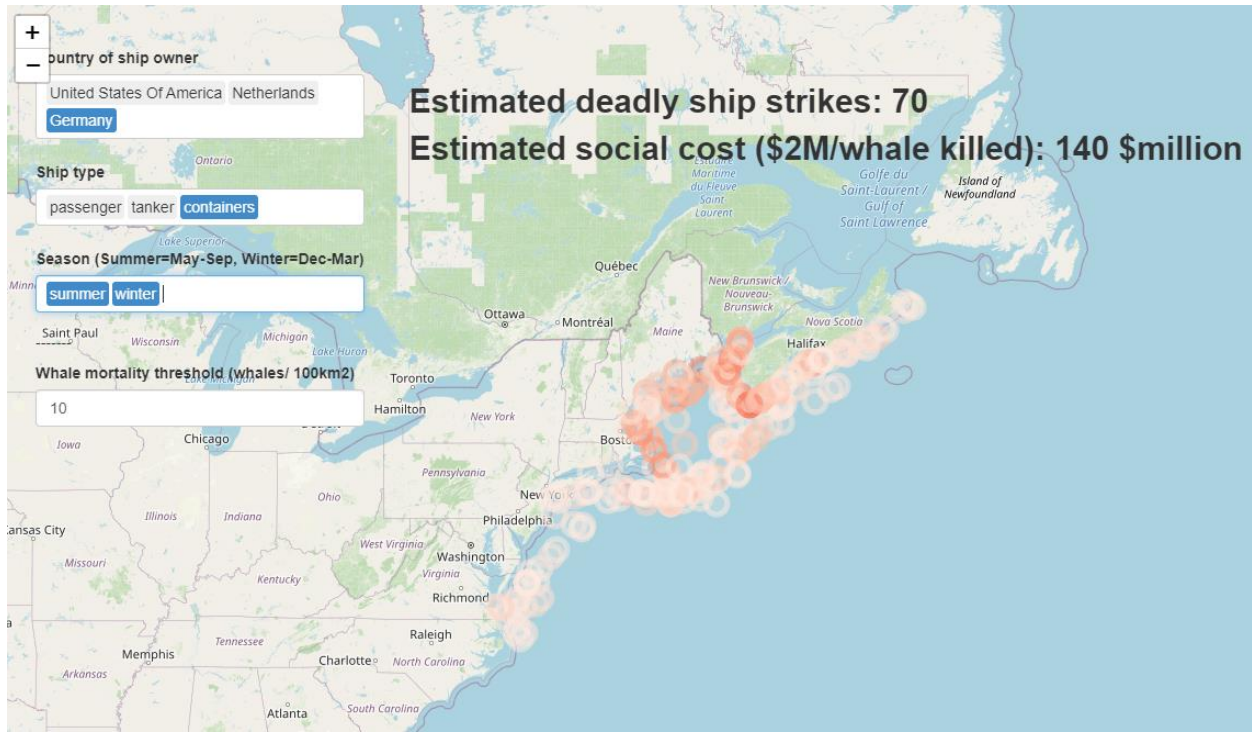
3rd Overall Outstanding Team: Multiverse of Data (Thailand): **Rice value chain**

Aiming to tackle SDG2 on zero hunger, the team utilized various big datasets, satellite imagery, and AIS to monitor the rice value chain from production to distribution, respectively. First, the satellite data was used to predict rice production in real-time by quantifying vegetation greenness and rice paddies. Then, the machine learning models were trained to predict rice production each year. Second, linear probability and logistic models were utilized to estimate rice exports (Thailand is a major rice exporter). And finally, AIS data was used to track shipments originating from Thailand and then benchmarked with the customs data.

UN PET Lab Hackathon

The UN PET Lab Hackathon was the first public data science competition held by the PET Lab. The goal of the competition was to cultivate the interest of data scientists in modern data privacy technologies, learn from user behaviors and educate participants in the core concepts behind PETs.

Data science competitions have been well known to engage this community, with hundreds of thousands of participants around the world competing on platforms like Kaggle, DrivenData, and Omdena to name but a few. These platforms break a given dataset into four, training and testing data, each having respective inputs and outputs. The goal from the perspective of the participant is to infer the relationships between input and output data using the training dataset and then to apply it to the test data to "guess" the corresponding outputs. Participants are then ranked based on a performance metric such as accuracy or precision.

This ranking element to the competition gamifies the learning process, giving the participants a feedback cycle to encourage their further development.

The concept of the UN PET Lab Hackathon was to mimic this process, but to add the caveat that all data exploration and inference had to be performed using privacy enhancing technologies. Specifically, the participants used Jupyter Notebooks on their computers to connect to a trusted execution environment (secure enclave) and the set of queries they were allowed to make were differentially private. Without going into the nuances of differential privacy, this is an output privacy technique (or data disclosure technique) that quantifies how sensitive an output is based on a set of parameters. These parameters, referred to as the privacy loss, are additive in nature and the larger the number, the more sensitive the result of the query is. The competition ranked participants based on their overall accuracy less the privacy loss.

The competition was conceived of, designed and developed in the six weeks prior to the dates of the 2022 UN Big Data Hackathon, which left little time to promote the event. Nevertheless, it showcased the power of the UN PET Lab network as many NSOs and universities from around the world shared the news of the event, resulting in 196 teams signing up to join the event, representing 328 individuals from 35 countries. There was also collaboration by a number of academic and private sectors to make the tools available to the participants. This included software from Oblivious, OpenDP (Harvard), IBM Research Europe and Microsoft.

The data considered in the competition was real data provided by the UNHCR pertaining to the effects of COVID-19 on refugees. The data disclosure team at UNHCR helped in the selection and preparation of the dataset and presentation of it to the participants. Federico Sanson from the UNHCR data disclosure team presented the data and Pauline Maury-Laribiere from FSO, Switzerland converted the original dataset into a format that would be engaging for the competition.

Further to the virtual attendance, there were three international on-site locations where participants could join from: Harvard University, USA; Toronto Metropolitan University, Canada; ETH Zurich, Switzerland. These on-site locations allowed participants to meet other like-minded individuals and learn from mentors who were willing to help them when they needed to.

Prior to the event's commencement, two webinars were run to introduce the challenge, the background theory and the tools which the participants would engage with. These webinars were 2-hours and 2.5-

hours respectively and brought together an international set of speakers from all of the collaborating entities.

Overall, the winning teams well represented the composition of the contestants with the top three teams coming from academia (University of Oxford), industry (Integrate AI) and the NSOs (the ONS, UK).

More information can be found on the website at: https://petlab.officialstatistics.org/

Annex 1 – Keynote presentation of Dr. Hamza Malik

Hamza Ali Malik, Director of macro-economic policy and financing for development division, UNESCAP

Thank you for giving me this opportunity to come and share and present what I call a narrative on what's going on in the global economy these days; what kind of near-term challenges we are facing and how to overcome them.

I'm also grateful to the statistics community of Indonesia for co-hosting this major International Conference on the role of big data in general. I'm an economist by profession and I like to think of myself as a storyteller; and economists cannot tell stories without information or without the scriptwriters; and the data community or statisticians basically are scriptwriters for the policy makers and economists to tell the narratives; and we need to check each other in these endeavors.

I'm going to focus on recent crises and economic prospects. The two crises in particular I'm referring to are essentially the Covid-19 pandemic and the war in Ukraine and its implications for the economic prospects, but I'll occasionally zoom out as well to talk about broader issues and challenges which we should focus on, which at times are neglected.

Let me start with the basic fundamental messages, the five key areas that I'm going to touch upon today:
(1) Primarily we economists have pursued, what I call the GDP centric development approach, which has come at a huge cost to social inclusiveness and environmental sustainability; so, there is a need to actually move away from this GDP centric policy making and invest in people and the planet.
(2) Then I'll talk a little bit about what's happening nowadays in the economies in the aftermath of the war in Ukraine and the pandemic; what kind of challenges are being faced.
(3) The policy constraints, which the global policy community are facing; fiscal positions have been constrained; debt is on the rise; inflation is high; interests are increasing. This is not the kind of situation policy makers normally envy for (or aspire for) to be operating in.
(4) The need to move away of what we call transition of economies from growth-centric models to more resilient inclusive and sustainable pathways and I'll show you some formal research; work that me and my team have done at ESCAP and
(5) lastly, I'll talk about the policy implications, and I'll talk about what kind of policy options are available. I may not go in the details of that because my colleague Rachel will actually pick up from there and we'll have a very nice engaging panel discussion with my fellow colleagues on the subject.

So, let's start with the big picture: the trend of GDP, the GDP per capita, employment and poverty trends since 1980. You can see the GDP has increased by more than three and a half times. GDP per capita, which is a proxy for income, has increased by about two and a half times since 1980. Employment has been increasing but by much less so and importantly the poverty has actually reduced by about three-fourths since 1990. So, this is the framework economists like to think about and that is why we tend to focus on GDP because we claim a high and rapid GDP growth is good. It leads to employment generation, it leads to income generation, it reduces poverty. So, if we focus on growth, then the rest of the things can actually take care of themselves. I tend to disagree a little bit. I will show you why focusing on economic growth has come at a significant cost to societies.

## Income distribution in Asia-Pacific
1980 = 100
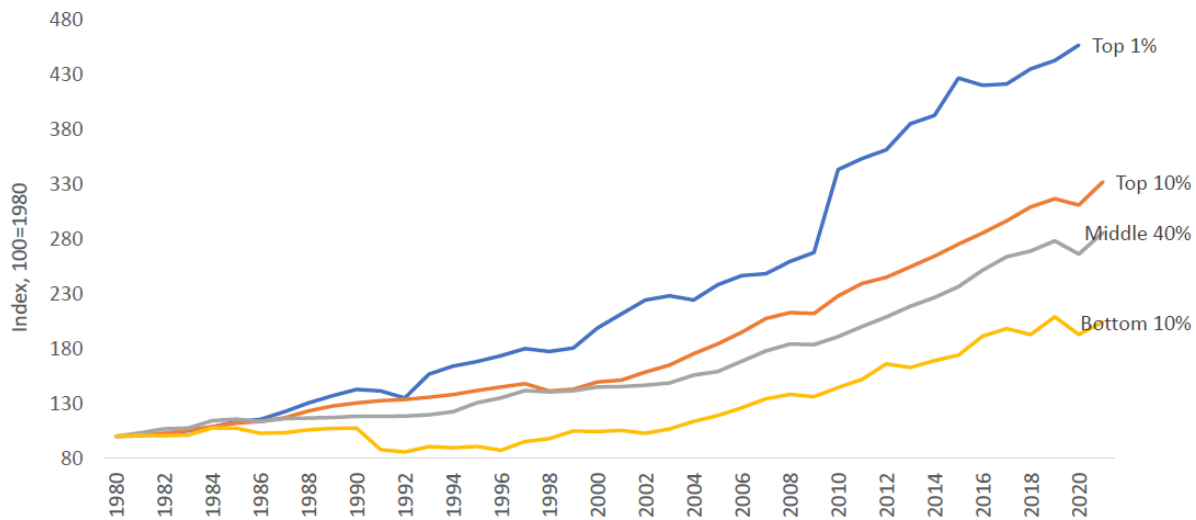
Source: World Inequality database (accessed on 3 November 2022).

This graph shows the data for the distribution of income; you could see the trends in different segments of societies in terms of their GDP per capita or income growth. So, you can see the top one percent have actually gained the most because of the rapid growth of last 30 years. This is the data for Asia and the Pacific, which I'm showing you, then comes the top 10 percent, then comes the middle 40 percent and then comes the bottom 10. So, the gains from rapid economic growth have been uneven in the region as a whole. Not everybody has benefited from economic growth. There's a fundamental point we keep highlighting in terms of making sure that economies grow by paying attention to inclusiveness.

Similarly, there have been significant impact on what we call environment degradation or considerable environmental impact of this rapid GDP growth. I am showing two indicators here: domestic material consumption and greenhouse gas emissions; they have been increasing in line with the rapid growth in GDP. The higher GDP growth has come with the cost both to the society and both to the environment but in the policy making world as economists we still tend to primarily focus on GDP growth alone and the understanding tends to be; let's grow first and then we will deal with inequalities later; let's grow first and then we will take care of the environment regulation later.

What I'm suggesting here is they need to change that narrative and a narrative change requires a good script and that's where the power of data and Big Data comes in handy, because it can lay out all sorts of information for you, for the use of policy makers, how to actually use that data to pay attention to the things that should be the objective of policy making as well; not as afterthought but as a priority. This is for me an essence of the sustainable development agenda, the three broad categories of the sustainable development goals: economic dimension, social inclusiveness and environmental sustainability.

Annual CO$_2$ emissions in billion tonnes (GT)
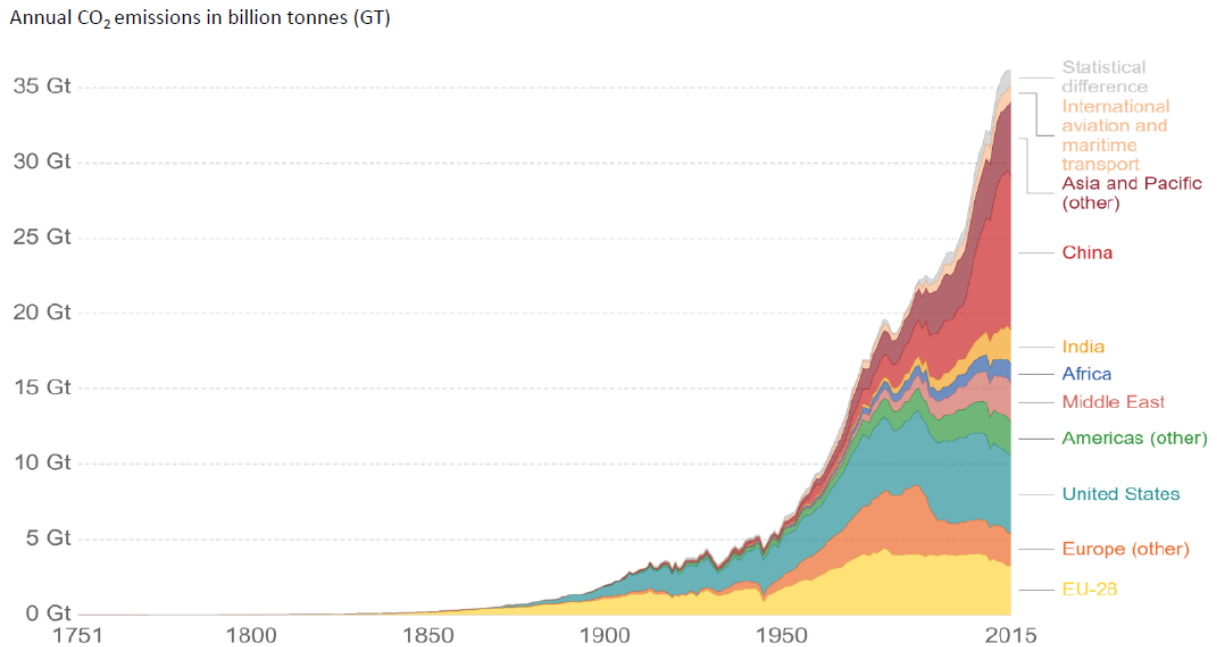
Source: Carbon Dioxide Information Analysis Centre (CDIAC).

If we zoom out a bit more, this graph shows the developments going back to 1751; you can see the damage to environment started taking place from about 1850 onwards. This is not a random year; this is the beginning of the Industrial Revolution. You will hear a lot of conversations in the global economic community now about climate change and climate justice; who's going to pay for the costs of the aftermath of disasters happening because of human induced environmental degradation; so initially you can see almost the entire 20th century the bulk of contribution to environmental regulation coming from US and Europe. They started to tone down their contribution only in 1990 onwards, very recently, and that's around the same time when the rest of the world took off. China in particular, India and many other Asia Pacific countries including Indonesia and all that is now a source of major debate; right, because the advanced economies like US and Europe are saying let's reduce our footprint on climate change; we need to do XYZ policies and the rest of the world is saying: no you have actually made a mess for the last hundred years or more; then it's our time now and we need to reuse the resources to be able to grow fast and reduce poverty and so forth; and that's a source of tension now because now the major contribution is coming from non-US, non-European economies. So, this is one example of how to present data and lay out a narrative, but it's up to the policy makers to interpret that.

After the big picture let me zoom in a little bit on recent crisis what's going on. This is the global GDP trend. I won't go into too much detail. The left-hand panel is simply showing the quarterly GDP growth. You can easily see that we had the global economy had a major dip beginning in March 2020 and by June the entire global economy was in a recession, albeit there are big differences between different regions. Latin America and Caribbean suffered the most; Asia-pacific relatively less but by and large the global economy went in recession by mid-2020. Then there was a rapid recovery that started to take place and by September 2021 things were looking okay when the Delta variant of Covid came along and then there was massive reintroduction of lockdowns and restrictions and so forth and the global economy started to slow down; then came the war in Ukraine at the end February of 2022 and it's still ongoing. We don't have real-time data a lot; we only have quarterly data available; so, you could see data for March and June, and you can see the downward trend beginning to take shape.

What do we use in economics when we look for high-frequency data or indicators, something that can tell me a little bit more, because policy makers are looking for evidence as a basis to take economic decisions. Everybody understands that the war in Ukraine is going to have many impacts on the economy. So, we use two proxy indicators: consumer confidence and manufacturing Purchasing Managers Index (PMI) to gauge consumer behavior and to gauge the production behavior in economies; and again, the visuals are pretty clear. You can see the consumer confidence has been heading south quite consistently across the globe and the same is the case for manufacturing PMI. These two indicators tell us a little bit in advance that the next quarter by the time real GDP data will come, we will probably see a little bit further decline, more contraction; so the global economy is heading towards further economic slowdown, possibly a recession further down the line as well.

Let's zoom in on our region: Asia and the Pacific; very similar story: the big dip in Mid-2020, big recovery after, but again there's a big difference between sub-regions. So, attention to details matters. If you look closely to the last part of it. Between March and June, you can see a big divergence happening in different sub-regions, in different countries. Some countries picked up; this is South and Southwest Asia primarily thanks to India; Central Asia (the orange line) is actually heading south. This is what economists call sometimes a k-shaped recovery: some economies are doing better while some others are doing worse; and that creates divergence between economic performance across the countries and creates many other tensions with policy making. But again, these are early days; only one quarter of data that we have to actually show that says yes, the global economy is heading south but some are doing better.
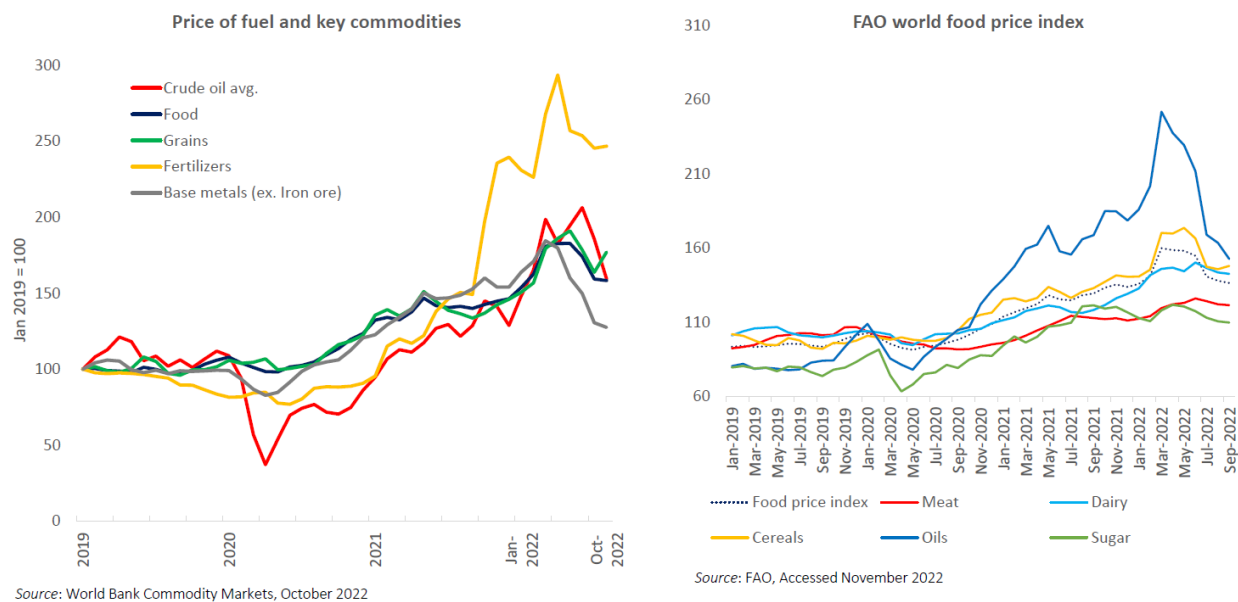
How do we know we're going to go forward? How to make a forecast? We can look at within the region or within the economies as well by different sectors: like hospitality, transport, wholesale and retail, manufacturing. So, basically those sectors that had more exposure of human interactions suffered the most (not surprisingly) because of the pandemic. Contractions big time. They haven't recovered fully yet and they're operating below the pre-Covid levels. So, the simple story is: yes, there is recession but it's uneven, both within and between countries.

It's not just GDP dynamics that we should pay attention to; I just want to show you the impact on what I call people and livelihoods. So, the left-hand panel is the impact on jobs employment; we're using the data from ILO in 2020 in Asia Pacific. As a whole 140 million full-time jobs were lost; some recovery was there but in 2021, the orange bar, about 65 million jobs, were still lost; so, there are major losses in livelihoods essentially across the board. The right-hand panel is based on our research in ESCAP in my team; impact on Poverty of all these economic dynamics that I was showing; and again, you can see using the metric of one dollar ninety that (on the blue bar) 85 million people have been pushed back into extreme poverty; if you use a different mark the impact on Poverty of economic slowdown is fairly considerable. The pandemic came with a heavy cost to people and their livelihoods.

This is something important I'm going to come back to later in terms of fundamental messages: the importance of resilience. Shocks leave long-term scars; it's not about one quarter GDP up and one quarter GDP down; you need to pay attention to what happens over time. So, the pandemic is likely to leave long-term impacts or what we call long-term scars; labor markets and people skills have been eroded and the chances to re-enter in the labor market have actually come down in many economies; there is a significant reduction in capital Investments which affects the future productivity of economies and most importantly there are learning deficits that are emerging now, because a lot of children across the region could not attend schools because they did not have access to ICT connectivity; and this

aggregate impact is costing about two trillion dollars of output, which has been lost because of the pandemic in 2020 and 2021.

The things were barely stabilizing and then comes the war in Ukraine. I'll be very brief here. There are three broad ways to look at the impact of the war in Ukraine on economies: first one is through the impact on prices: food prices, commodity prices, like oil or fertilizers, and so on and so forth; second one is through the weakening of global demand; our region depends big time on our exports; so demand from outside of a region has been affected; and third one is a sheer amount of uncertainty which affects investors and consumers; people are nervous to take a long-term position; the investment decisions require this kind of clarity and then there's all sorts of details; so rather than explaining these one by one, I'll show you some visuals of how these are actually affecting the policy making and the economic decisions. I've tried to make it very visual, as I said. I'm trying to tell you a story here based on the data that we use and how we use data. Data is not neutral. I'll give you an example in a couple of minutes.

Price of fuel and key commodities

Crude oil avg.
Food
Grains
Fertilizers
Base metals (ex. Iron ore)

Source: World Bank Commodity Markets, October 2022

FAO world food price index

Food price index — Meat — Dairy
Cereals — Oils — Sugar

Source: FAO, Accessed November 2022

So, this is the trend in commodity prices; across the board prices significant increase; again, for the ease of comparison I've normalized them to 100 and you can easily see most of the commodity prices have actually gone up; one small point of attention: the rise in prices started before the war in Ukraine; it started mid-2021; so in a lot of economic commentary you will hear that the war in Ukraine led to inflationary pressures; that's not true; the inflationary pressures were picking up before that and why: because there was pickup in demand that was coming back on track after the restrictions were eased in many economies; further, during the pandemic, economies across the globe introduced extremely extensive expansionary policy packages through various monetary and fiscal measures (about 1.8 trillion dollars of worth of liquidity), which was injected in the system and that did not disappear in thin air; it's going to fade into some pressure on demand eventually and that was showing up at the beginning of 2021.

I remember having very interesting and heated conversations with our colleagues in the Regional Commission, with IMF and other colleagues, and they thought that this is not going to go up, whereas we in ESCAP took the very clear position that this inflation trend is going to go even higher; primarily because of this pickup in demand and very expansionary policies that countries thought they had to

introduce but they should have been ready for the consequence of those things; and then came the war in Ukraine to add insult to injury in many ways; and that is why you see huge prices. Look at the yellow line for instance that's the price of fertilizers; higher prices of fertilizer mean that the basic input in agriculture commodities is more expensive; so food prices are likely to remain elevated for some time; the right hand panel clearly shows that as well; significant impact also happened on oil because of the disruption of one of the major exporters of it (the Russian Federation).

This is the global inflation picture. You can see that the red dots are essentially inflation targets, while the blue lines are the actual inflation; the headline or core inflation are simple visuals, which tells you that inflation in advanced economies is significantly higher compared to their stated targets and objectives; consequently, the interest rates are on the rise. The interest rates were in the negative territory for a long time, but now they have started to go up and started to go up very drastically, especially in the US which is the red line here; so they have recently increased the interest rates again. In fact, this means that policy makers in Asia Pacific cannot just say we don't want to change interest rates, just because US or United Kingdom are increasing interest rates.

You cannot avoid increasing interest rates because if you don't, there's going to be capital outflow happening from the region; because it's then more lucrative to invest in US compared to our region. So, there is pressure on economies in the region to increase interest rates; if they don't, they're going to experience exchange depreciations which in turn feeds back into inflation trends and so on and so forth; so, it exacerbates the challenge. The picture for Asia Pacific shows a very similar trend. Things started to pick up by mid-2021 before the war in Ukraine; and since then, we have been seeing a rising trend across the board exacerbated by the war in Ukraine. Same thing on the right-hand side, you can see inflation has been significantly higher compared to the targets, which are the blue triangles, in almost all economies; some economies are suffering the most like Turkey, Sri Lanka, Iran, Lao PDR. Inflation is massively high in those countries, creating serious problems for policy makers for people and the overall economic prospects.

The graphs below show how central banks are responding. So, monetary tightening increasing interest rates in advanced economies is one factor, but domestic inflationary pressures is a second factor, which

Monetary tightening in advanced economies and domestic inflationary pressures are leading to rising interest rates in Asia-Pacific…
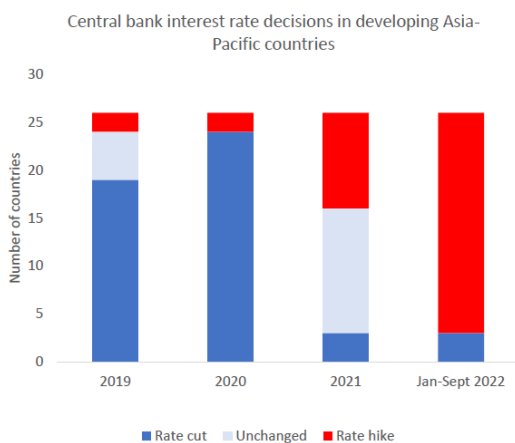
Source: Based on CEIC (accessed 21 October 2022).
Note: Central bank interest rate decisions are based on policy rate data for 29 countries in developing countries in the Asia Pacific region
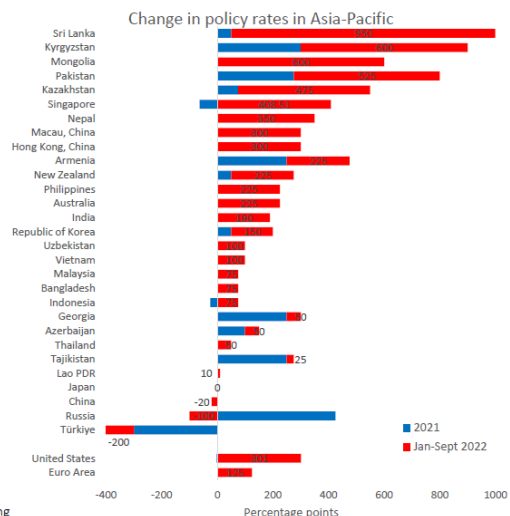
is causing the policy makers in the region to increase interest rates as well; and the whole scenario has changed very quickly, very drastically. In the left-hand panel the blue bar indicates countries that reduced the interest rates, the light blue bar is policy maker that kept the rate unchanged, whereas the red bar indicates increase in interest rates; see how the picture changed between 2019 to 2020 when almost all countries were reducing interest rates and how we changed the situation in 2021 and then more so in 2022; most countries are now increasing interest rates. In my own assessment they are behind the curve. Further interest rates increases are likely to come. The right-hand graph gives country level detail in Asia Pacific. You can see almost all of them have increased the interest rates in 2021 and 2022. Our assessment is that interest rates are probably going to be even higher in the coming months and years.

So, the dollar is becoming stronger, as I was saying, because relatively speaking the interest rate increase is much more pronounced in the US and other advanced economies compared to the region. Approximately 75 to 80 billion dollars have already flown out of our region to those kinds of destinations, because it is more lucrative to go there. Money always chases returns. So, the US dollar by consequence has gained strength and countries in Asia Pacific are experiencing exchange depreciations. People don't want to invest in their currency that's why the value of that currency is declining most pronounced in Sri Lanka, Laos, and Turkey. These are the same countries that are experiencing high inflation, because inflation and exchange rates are not independent variables in some sense.

So, below is one of the most important graphs in a way that I'll show you. How to interpret the data here? Just to tease you a little bit; so, the fiscal space has dwindled; financing pressure is high; debt is up and fiscal consolidations are expected. What does that mean? First of all, look at Asia Pacific; it was fairly stable on average when it comes to fiscal deficits up until 2019. It was hovering around one percent deficit which is very small compared to global standards; then came the pandemic; massive injections for fiscal measures which were much needed. However, consequently the deficit went from on average about one percent to about five percent; and then the debt went up at the same time; so, the deficit is

Fiscal space has dwindled, debt is up, financing pressure is high, and fiscal consolidations are expected …
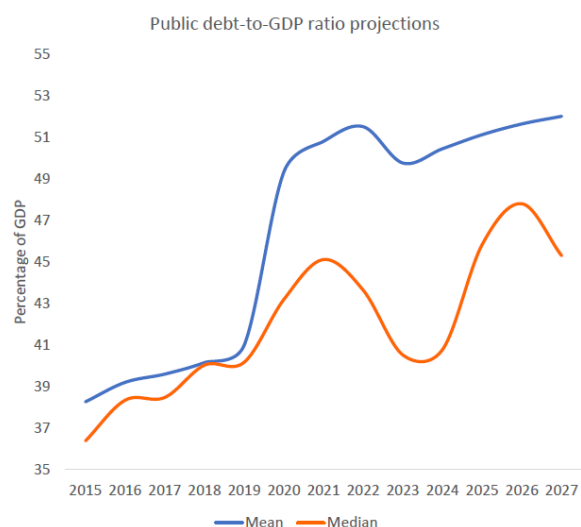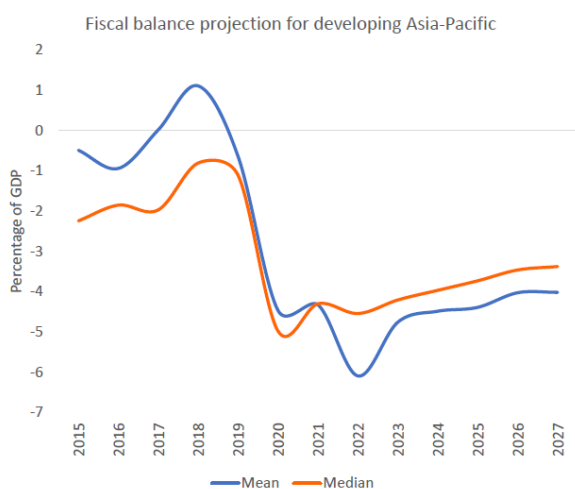
Source: IMF, Fiscal Monitor Database, October 2022.

going up; debt is going up; which means the need to finance that gap is increasing. You need more money to close that gap, right; so, where is the money going to come from; well money is already

heading outside of the region; you're not attracting enough capital flows; so, there's significant liquidity and financing pressure being faced; as a consequence, the price of that liquidity interest rate is going up; it's becoming very difficult now to find money.

Now here is the twist in the narrative for our economies. This IMF graph shows the forecast for the future of 2022 onwards; why is the graph not showing that the deficits are going to increase? Some observers will say: well, the countries are taking measures to stabilize fiscal deficit around five percent; we talk about physical consolidation; cuts in spending; increases in revenues or taxes to try and close the gap, as you will.

We are not supportive of that approach; it builds in significant contractions, like cuts in spending on social protection, on health care and many other measures, where we strongly believe that they're important for sustainable development goals. So, these are in a way the desires of IMF, which they are suggesting to countries. These are not neutral projections. That's the power of data.

I can start from the same data and then make different projections and provide a very different narrative, which is going to be saying that I want countries to continue to spend on essential needs; yes, they will run high deficits; yes, the debt will go up, but let's talk about how to finance that rather than approve cuts; it's a whole different policy prescription coming from the same data, right; so that's one example. I'll share one or two more examples like that to show how you can interpret the data differently.

Based on our research I will show you one graph, which shows the importance of fiscal policy or fiscal measure; or why fiscal consolidations are actually not good; well, fiscal consolidation tends to lead to increases in inequalities; research firmly confirms that; and then it reduces the income of societies for some years; that is why we are not in favor of fiscal consolidations, although from IMF perspective, the way they look at situations, they prescribe something slightly different

Fiscal measures do support inclusive development … but impact varies ESCAP

Impact of seemingly same category fiscal policies <u>varies greatly</u> raising questions on effectiveness and targeting
Impact of fiscal policies on inequality, according to 1 per cent of GDP expenditure/revenue

Gini coefficient (0-100)
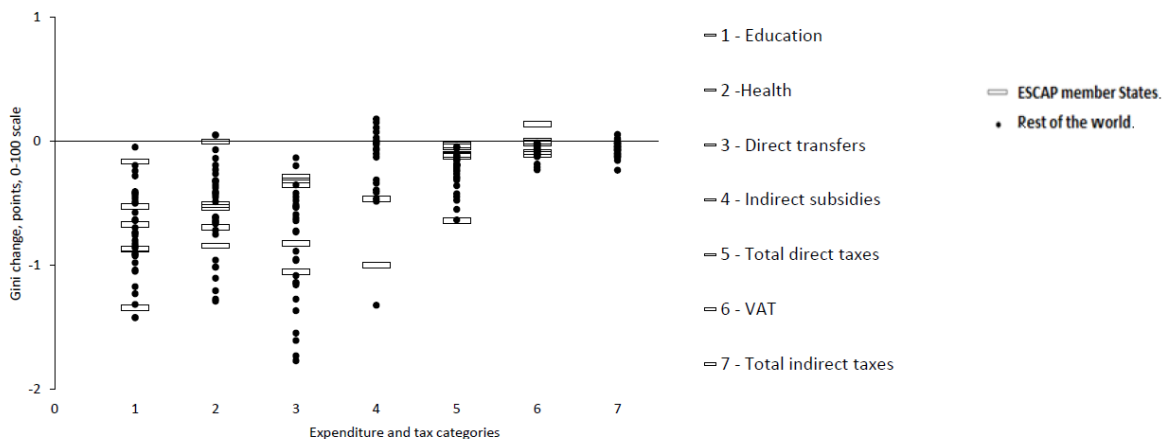
Source: ESCAP estimates based on CEQ data.

This graph above took us four months to put together, or more actually. This combines all the fiscal measures of Asia Pacific economies and the rest of the world in one visual. What we simply wanted to do was to isolate which fiscal measures are most effective. So, there are examples of investments spending side on Education, Health and direct transfers; and then there are other examples like taxes, value added tax and other indirect taxes; and we assessed the impact of these policies on average across the region. So, the white bars are essentially the Asia Pacific countries, the other dots are other advanced countries depending on the availability of data. Again, you can see that the impact of spending on education and health and direct targeted transfers is much more pronounced; they are much more effective in reducing inequalities as compared to taxes. So, the lesson we draw from this very extensive research is that targeted expansionary fiscal position in education, health and social protection helps reduce inequalities. Thus, our recommendation is: countries should try and find means to continue spending in these areas even if the deficit is high. We don't need to institute fiscal consolidation early on because it has very negative consequences, especially from the inequality aspect.

This is another slide that took a long time (four or five months). The left-hand panel shows the Asia Pacific risk landscape; so, we looked at 462 episodes of shocks since 1960 for the Asia Pacific region and we categorized them in four different categories: financial crises, terms of trade shocks, pandemics and natural disasters. We simply wanted to isolate what kind of policies work under different circumstances; so, we can draw lessons from pandemics and from different kinds of shocks. You can see that the Asia-Pacific region have been experiencing many more shocks since 1980s compared to the period before, because the region is growing and getting more integrated and so on so forth; and then we looked at these policies and see 60 years later, which policies have worked in the past and why. So we can prioritize those policy areas as we build back better.

So, these are just examples of mainstream economic policies. I'll show you one or two visuals only. You need to pay attention to these shocks because some of them tend to have long-term impacts. The first one is the impact on investment; the financial crisis caused significant slowdown in investment for several years, while pandemics lead to increases in inequalities for several years. That is why I was emphasizing earlier that you need to pay attention to reduce inequalities through fiscal measures because it's not a one-off thing; it's not going to go away in a quarter or so; it's going to leave a long-term impact for several years. The same is the case with natural disasters; they tend to create havoc.


… and pandemics have more severe socio-economic impacts … in countries with pre-existing structural vulnerabilities
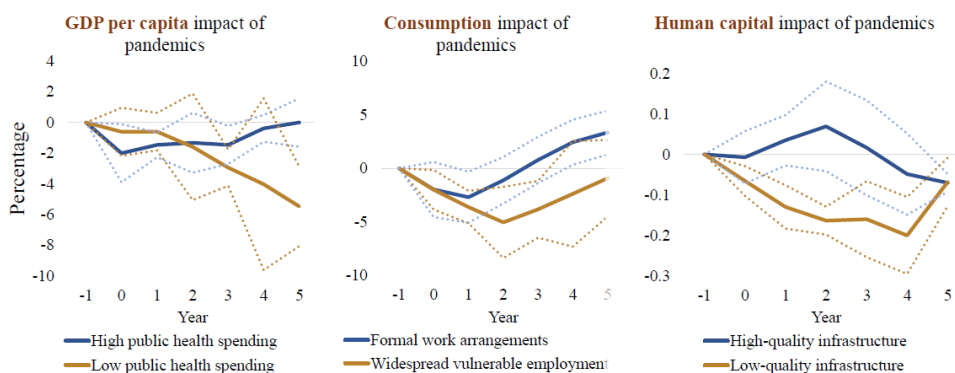
Source: ESCAP estimates and forecasts.

This slide above shows that the socio-economic impact of pandemics tends to be quite severe. You can see three examples here: one is high public health spending versus low public health spending; similarly for formal work arrangements versus widespread informal work arrangements; and finally, the quality of infrastructure. You can see the performance actually being different. These blue lines are what I call resilience enhancing policies whereas the orange lines indicate less resilient enhancing policies. When we isolate the performance of these economies, you can build evidence why certain policies matter for certain cases.

Difference in resilience enhancing policies explains difference in development performance
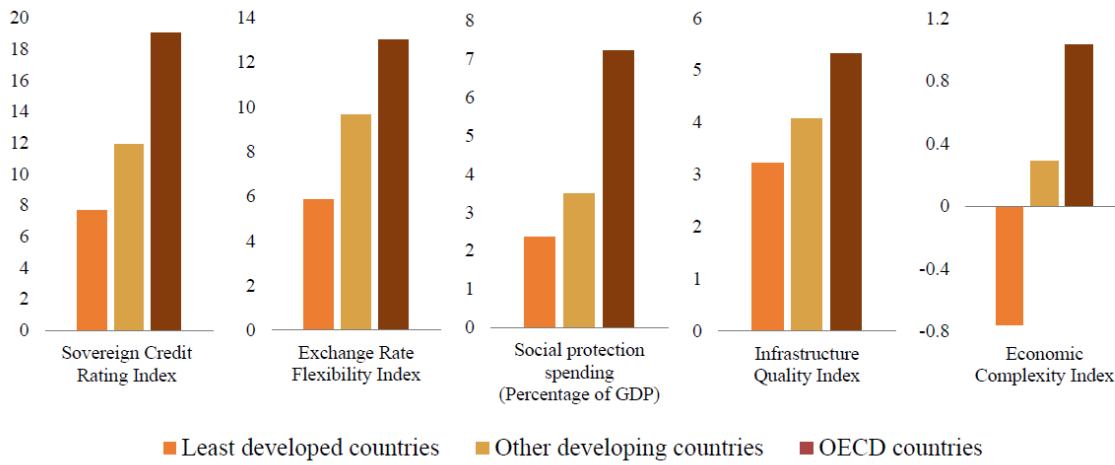
This is just one visual, and you can find in our research several other examples of sovereign flexibility investment, social protection, infrastructure quality, and economic diversification. There are many indicators which we looked at for the last 60 years for each specific countries, and then the visual is very clear: those countries that invest in resilience tend to have better development performance; in other words, it pays off to reduce the impact of setbacks compared to only prioritizing rapid growth, because you want to protect the development gains you already had rather than mindlessly pursuing more and more expansion all the time; because once the shock hits, it puts you back several years, and the frequency of crises has been increasing in the region quite a bit.

A transformation of economies towards resilient, inclusive, and green development pathways is needed
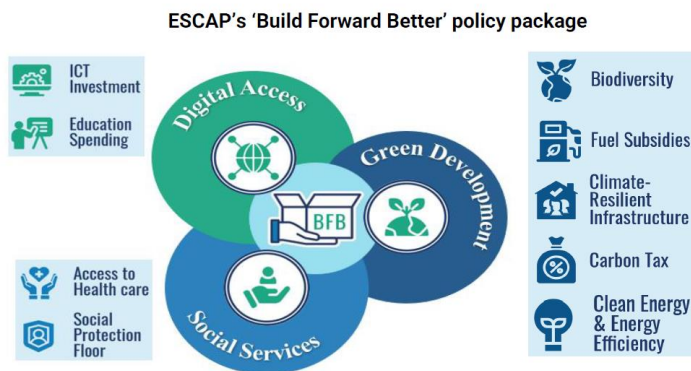
Based on this research we proposed and developed a macro-econometric model, which we have actually applied with Bapanas in Indonesia. This is a three-pronged strategy what we call "build forward better" policy package. It pays attention to Digital Access, Social Services and Green Development; and the model can be

adjusted depending on the country's own priorities and within each of the three broad categories we have several areas that can be quantitatively explored what happens if you invest in this area versus this area versus that area and the implications.

Such transformation offers notable socio-economic and environmental benefits … if ambitious investments are undertaken …
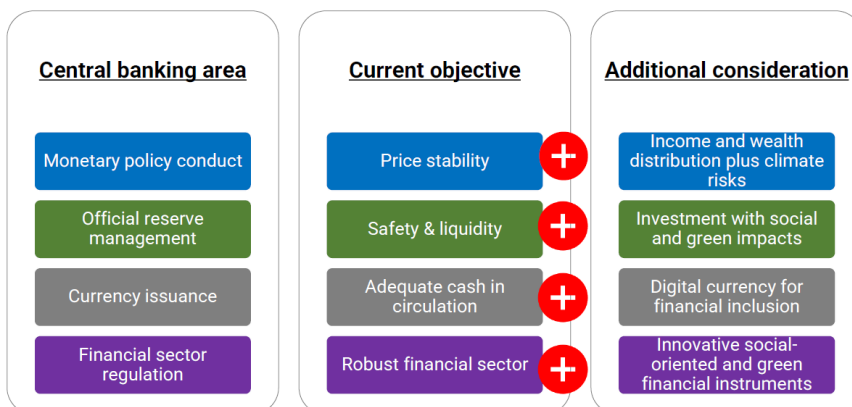
Here is one visual of that for the Asia Pacific region. The blue lines indicate the business-as-usual spending, which is the simple continuation of current trends; and the dark lines indicate the situation if countries actually invest the way we are recommending in those areas. You can see the beneficial impact on reducing poverty, reducing inequalities, and reducing carbon emissions; so what we are trying to build is evidence that it is possible to simultaneously have economic growth, social inclusiveness and environmental sustainability provided that we have a little bit long-term horizon of about 10-12 years; and then issue policies that invest in resilience and so on so forth; and you end up having better outcomes over time. Obviously, investment of this scale requires a lot of resources, which means the debt is likely to go up. I've already made that point. It's easy to say invest in this area; or invest in that area. The million-dollar question is: where is the money going to come from?

On the policies, I'll be deliberately quick, because as I said we'll probably pick up some of these issues in the discussion in the panel session; so, on inflation we have two or three broad areas: one is what I call timely increase in interest rates; timely is important because if you delay it, then you're going to experience capital outflows and exchange depreciation having a feedback effect on inflation. Somebody can easily ask a question: well, how can increase in interest rates bring prices of food down? People don't see the direct impact. There are many channels through which food prices can be affected. You can cut consumption tax and import tariffs to have more availability of food items, which will reduce the prices; and there are the more direct administrative steps that can be taken, which I can elaborate on during the panel discussion

A new proposal in our research at ESCAP is to push the boundaries of Central Banking. I'm an ex-Central Banker myself, so I understand the dynamics of monetary policy makers who only focus on targeting

inflation. What we want is inclusive Central Banking; so Central Bank should go beyond just arguing inflation, they need to think innovatively and outside the box.

Beyond inflation … inclusive and green central banking   ⊕ **ESCAP**
                                                                                    MOVING FORWARD TOGETHER

Central banking area	Current objective	Additional consideration
Monetary policy conduct	Price stability ➕	Income and wealth distribution plus climate risks
Official reserve management	Safety & liquidity ➕	Investment with social and green impacts
Currency issuance	Adequate cash in circulation ➕	Digital currency for financial inclusion
Financial sector regulation	Robust financial sector ➕	Innovative social-oriented and green financial instruments

Central Banks can actually pay attention to inequalities and climate risks integrated in the decision-making process; and then take decisions; they can actually use the massive reserves and channel them for social and green investments in the economies. Central Banks can introduce digital currencies, which can actually have a massive impact (according to our research) on financial inclusion and lastly, they can use their regulatory power to encourage investments to flow in those areas that promote green development.

Regarding fiscal matters, I'm dividing these into two categories: one is the domestic measures and the other is what is needed by the international community or by multilateral approaches to try and create fiscal space. The crux of the message I am conveying is that, yes, you have policy options but implementing them requires fiscal resources; so here are some ideas on how to create that fiscal space at the domestic level. These are the immediate and short-term measures that you can take. As for the multilateral considerations, these are the kinds of discussions that are taking place in the UN led by the Secretary General himself. I'm part of one of these teams; so, for instance, extended debt services suspension initiatives can help to give a country a little bit more space to channel the resources for development priorities; and on the next slide there are some of the medium to long term measures countries can consider in increasing fiscal space which is going to be the essence of trying to effectively pursue the 2030 agenda.

Key Takeaways   ⊕ **ESCAP**
                                                                MOVING FORWARD TOGETHER

- Focus on "reducing setbacks" (building resilience) is as important as "accelerating progress" (GDP growth) … invest in people and planet.

- Rethink and transform development pathways … focus on social services, digital access, and green development.

- Near-term economic prospects look grim … current challenges are likely to persist for next couple years.

- Room for supportive fiscal and monetary policies has become increasingly limited … several options are available to increase policy space.

So, I stop here. Here are a few things that I want you to remember. So, focus on reducing setbacks; in other words, building resilience is as important as pursuing rapid GDP growth, or accelerating growth, because shocks leave long-term scars; and you lose your development gains over time. How do you build resilience? You invest in

education, invest in health, and invest in social protection. I showed you evidence why they are beneficial for the economies. Secondly, we need to rethink not just economic thinking but the actual development pathways. So, you should deliberately and simultaneously pursue resilience, inclusiveness

and sustainability, not GDP growth first and then another consideration. The model that we have proposed for the "build forward better" policy package on three areas can actually do that.

The near-term situation is looking pretty grim to be honest. Things are going to get most likely a bit worse in 2023. The crux of the entire thing is: what are the policy options? what can policy makers do? Well, there are some options that have been very briefly presented both on the Central Banking side and on the fiscal side leaving hopefully enough interest in the audience.

Thank you very much for your attention and I look forward to engaging conversation in the panel.

Annex 2 – Keynote presentation of Dr. Carlo Cafiero

Dr Carlo Cafiero, senior statistician and economist, FAO

What I will be discussing today is where we are in terms of informing policies that are intended to improve access to food for people. Before going into the details, I want to make a small premise to highlight what we currently look at. "Food security" in a way has a different meaning than what is probably still commonly understood. By food security we essentially mean the outcomes at the individual level. So, when we discuss food security it is no longer talking just about food availability but rather about the individuals ability to access the food they need continuously and of the right quality.

The individual food security and the nutrition outcomes related to food is the result of a complex interaction of factors that happen at very different levels. Some factors that are more proximal to the people, some that are more distant, including phenomena that occur at the global level. So, this is the reason why we see food security as the result of many interactions. This is quite a departure from more traditional ways that see food security only as the adequacy of the total supply; so, you would just look at the food balance. It's more important to understand what happened with the food that is available and how different people can actually access it.

Another premise is that we define food security as composed of several dimensions and recently the definition has been expanded to include not only availability, access, utilization and stability, but also agency and sustainability to recognize the importance that people take control of what they eat of how they utilize the food.

Regarding the opportunities presented by Big Data, I do not need to spend many words because we have already been getting hints that Big Data can help in many areas. Given the complexity of the system, there are definitely many areas in which data is useful and can be effectively utilized to guide policy, and I won't go into specific details of examples of specific applications because I think in the course of the conference this will be covered by people that are much more competent than I am. Moreover, the hackathon might probably be an opportunity also to show additional utilization.

So, I will focus my remarks mostly on a few fundamental challenges that in my humble opinion still prevent the full potential that exists in new data. We need innovative ways of reducing data to effectively inform food security policy.

So, this is a point that was already raised in the previous discussion, because we are interested at outcomes at individual level what we need is to be able to detect inequalities while recognizing that we live in a very diverse world, and this is where granular data coupled with careful statistical analysis is necessary to distinguish and not to confound diversity, that we want to cherish and promote, with inequalities that we want to fight and oppose; and to do this we need a new model; we need innovative thinking that very stylized traditional microeconomic models are not able to capture, and reflect the diversity. We are not all equal and this is very good, but we all have a fundamental right to access the food that we need and so this is what we need to achieve.

The final premise, or initial caveat, is that because of this complexity, because the space in which data is generated and information is used, is so complex and so wide, it's very important to have integrated governance systems, including for the data, because decision making is not happening only at one level.

In the 2022 HLPE report on data collection and analysis tools for food security and nutrition, we introduced a data-informed decision-making cycle (see Figure below) to highlight the circular features of how data enters into decisions made at different levels; with a lot of feedbacks so that information produced by some can be used as data in a following steps, and all of this is only possible if we have an effective data governance and we'll connect.

FIGURE 2:
DATA-INFORMED DECISION-MAKING CYCLE
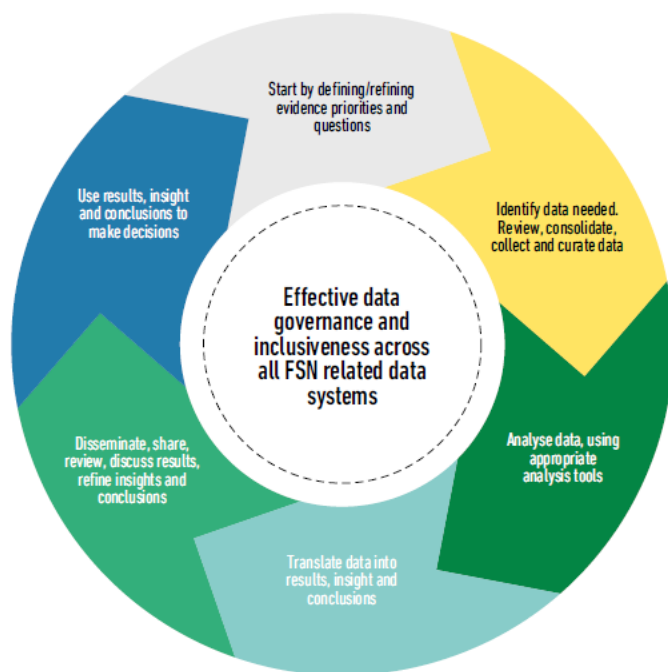
Source: Adapted from Piwoz et al., 2019.

It is clear, that there are plenty of opportunities for Big Data and I am focusing here on one specific area, which is food consumption data. So, what do people eat? This is a very important domain on which we still lack a lot of information and Big Data presents a lot of opportunities. Just to give you an example of the complexity.

Critical urgent data gaps areas

Level	Dimensions of food security and nutrition					
	Availability	Stability	Sustainability	Access	Utilization	Agency
Macro	Natural resource base Earth Observation International food commodity stocks and trade	Global/regional food commodity stocks and reserves	Weather and other risk trends and predictions	International food commodity prices	**Food composition data** Food safety data	
Meso	Domestic food availability	National food stocks and reserves		National food price indices	Water & Sanitation	Market concentration shares
Micro	**Local food systems**	**Early Warning Information Systems**	**?** (Environtmental, social and economic sustainability of farming)	**Local food prices Household incomes and consumption patterns Food insecurity experiences**	Household living conditions **Household water access Dietary intakes**	Food insecurity assessment surveys Women's Empowerment in Agriculture **Rural Livelihood and Information Systems**
Individual (Outcomes)	Dietary intake/diet quality; malnutrition prevalence and related health outcomes					

Again, this is a matrix that shows by level and by dimension of the food security concept, what kind of data we would like to have; and each of these cells is equally important. This is a picture of what's there and it's not a very pessimistic picture. There is a lot of information out there, even though sometimes it's fragmented; sometimes it's difficult to locate. There are also areas

where we still have fundamental gaps and I think that this is an area where we as a community can do a lot, and can also join forces with the private sector to provide the highest benefit.

We lack for example data on food composition, the composition of local foods. As I said, we live in a diverse world. The kind of food that people eat is very different and so if we want to evaluate and assess the quality of the diet from a nutritional point of view, we need to know what's in the food we eat and, strangely enough, this is still an area, in which we don't have readily usable information. The situation is more problematic in the last row of the matrix, which is the micro level. It's at the level of individual consumers and at the level of individual farmers. This is where we still rely on traditional data collection systems based on farm surveys as far as food production is concerned.

I'm very glad to be here in Indonesia, a country where there is one of the strongest and more effective traditions. For example, in collecting household survey data the Susenas (the national social economic survey) is probably the most developed household survey and it's conducted every year at a scale that is incomparable with what happens in other countries; and BPS does an incredible job in making the data available very quickly; but this is an exception; in most countries of the world and I say including developed countries, we lack timely granular information on what people eat and what people consume; and the same we could say on local food systems; how food production occurs at the very local level; so there is plenty of opportunities to be exploited.

In particular, as I said, we need granular information because individual characteristics like gender, age, education or employment status are very important, and they depend on the local context, and those characteristics may condition the ability of people to access food, and because of the challenges to collect this information through traditional household survey, we now hope to be able to use new technologies in order to obtain this information in a much more timely manner on personal devices like our cell phone, our participation to social networks, even the information that we provide when we order food for delivery etc. It's something that only a few years ago was simply impossible.

Another important aspect is that modern forms of food consumption are simply not captured in the tradition of household level-based survey; so, we have a hard time understanding food behaviors when people consume food away from home which is becoming a dominant way of getting food. There is a very useful and important attention given to the relevance of food for health, and hence the need for food quality data, and dietary quality data.

One point, that I want to make, is that the technology, that we have nowadays, potentially gives us the possibility of getting that information related to food consumption. This information could flow in both directions. So, from people to policy maker, and from policy decision to people. "Agency" as food security dimension means giving people more power in controlling what they eat; there is still lots of things that can be done to improve and increase our collective ability to benefit from data and therefore I want to just focus on a few that in my opinion are key challenges.

The first is that we still live in a world with a wide digital divide and overall insufficient data literacy. I think that it was mentioned already in the morning, that not everyone has the same ability to access the technology and to use it effectively; even among national statistical agencies the capacity of making full use of Big Data, of innovative new data, is not equal; there are big differences across different countries.

Yes, you can access this data and it's an ocean of new data, but do we have the analytical capacity to make full sense of it; it's like I heard this expression, and I like it a lot: we swim in an ocean of data but

we risk to drown out of ignorance, of how to make sense of this data; the reason is because we were in a way still trying to catch up with, what I would call, a revolution in knowledge and information making; but statistics people don't realize how young statistics is as a science. We still need to understand how to make sense of uncertainty, when we have been flooded with data; and then data science has come around, even before we have reached the point in which statistics is taught more widely. Students only start studying statistics when they reach college level (or in some cases high school); and this is a problem because most people in the world don't have the right frame of mind to reason with partial information and uncertain information.

One problem is the enormous diffusion of high computational power; you know that today we are carrying in our hand more computational power than the computer which send the man to the moon but do we really have the capacity to use it. There is an illusion that we are able to substitute the thinking with the artificial intelligence or the algorithmic solution which is an illusion because you still need to have proper human thinking to make sense of the information.

I think that we need to address this problem, because there is also a divide in developing this ability; and when there is a scarce resource, there is a possibility that rent seeking behavior starts to appear from those who possess this scarce resource. In this case I'm talking about analytical capacity. Some will try to use it for private personal interest as opposed to the public good, and I think that one of the problem is also the fact that in many circles data, especially digital data, are still considered as if they were tradable goods; so there is too much of an economic view of treating new digital data, as if these were tradable commodities and I think this is an aspect that we should think about. So, I am very provocative here in thinking that what we need to address is data governance; so how to make sure that the data is shared and not sold.

I have heard people speaking of things like data ownership, or data as a strategic asset; I also heard talk about data as a weapon and I personally don't like all this. I think that data is a fundamental resource, and it should be an open access resource because of the characteristic of digital data. Digital data sits in many places. If I use digital data, I'm not reducing the ability of other people to use it, so there is a fundamental non-rivalry and non-excludability embedded in the fact that through the internet virtually everybody can access a piece of data that sits somewhere which is not under the control of specific individuals or people or institutions.

We are witnessing all these example of strengthening data protection to try to address a conflict between equally important rights; the rights to be informed and the rights to preserve our own privacy; but I think that these ethical aspects, of why we should favor more open free sharing of data, have not been fully explored even though there are very good examples like this article by Floridi and Taddeo, which speaks about how we should rethink the ethics of dealing with data.

I think that we should overcome the view of trading data as economic assets simply because the economic condition of excludability and scarcity do not apply to data. I am puzzled but I heard a lot of data protection when, in reality, what we should protect is not the data; what we should protect is the right of people; is the privacy. One problem that certainly exists is that the jurisdiction to make sure that data governance can improve is beyond the national level; it's at the global level and so one of the main concerns to an effective data governance is how to address the legal aspect that are still dominating and that prevent data sharing beyond the national border; because from where I sit in an International organization like FAO, I'm all aware of how important it is that we have data from all over the world; so

it's not only important for Indonesia to know what agricultural systems do in this country, because some of the consequence derive from things happening elsewhere.

So, in the mentioned report we have some proposals on looking at alternative ways in which digital stored data can be seen as a public good and how it should be given Open Access, provided that it doesn't infringe on the fundamental rights. This leads me to my most important conclusion, namely that an optimal level of data sharing in this context can only be based on trust. It is not based on economic incentive and is not based on strategic thinking; it's as the human development report recently described, we need a new social contract between people and governments and private institutions to share the data.

In this 2022 HLPE report we make a proposal to consider creating a Data Trust for food security and nutrition data, which means creating an organization that assumes and receives the mandate to decide which data that are relevant for food security and nutrition could be treated as fully open, and also monitor and sanction possible abuses of such data; and also certifying that data is fair. I don't know whether or not such proposal will be taken on; but even before such an arrangement materializes, we need to invest into statistical and data analytics capacity development at all levels and implement mechanism that prevents rent-seeking behaviors and make sure that the benefit of this information goes back to the people for which the information is generated and used.

With that I thank you for your attention.

The 2022 HLPE report on *Data collection and analysis tools for food security and nutrition: towards enhancing effective, inclusive, evidence-informed decision making*, Committee on World Food Security
https://www.fao.org/3/nj986en/nj986en.pdf

Annex 3 – Keynote presentation of Prof. Dr. Bertrand Loison

Thank you very much.

I will speak this afternoon about the work that we have done in Switzerland regarding the creation of a data science competence center and about data strategy and data governance.

I will cover two topics the first one is data analytics and the second one is data governance, but, first of all, I will introduce the motivation why we created a data science competence center. Back in 2015, the message was quite clear: we are in the middle of a data revolution, and you need to anticipate the effect of this data revolution on your work, if you still want to be relevant in 20 years. In the UK they published a very interesting report asking the question: are we still measuring the right thing, if you want to understand our economy? We have technical challenges, and we have also data analytics challenges, and it was really the question of how can we, as official statisticians, react to this new world?

… and the Response of the Federal Administration
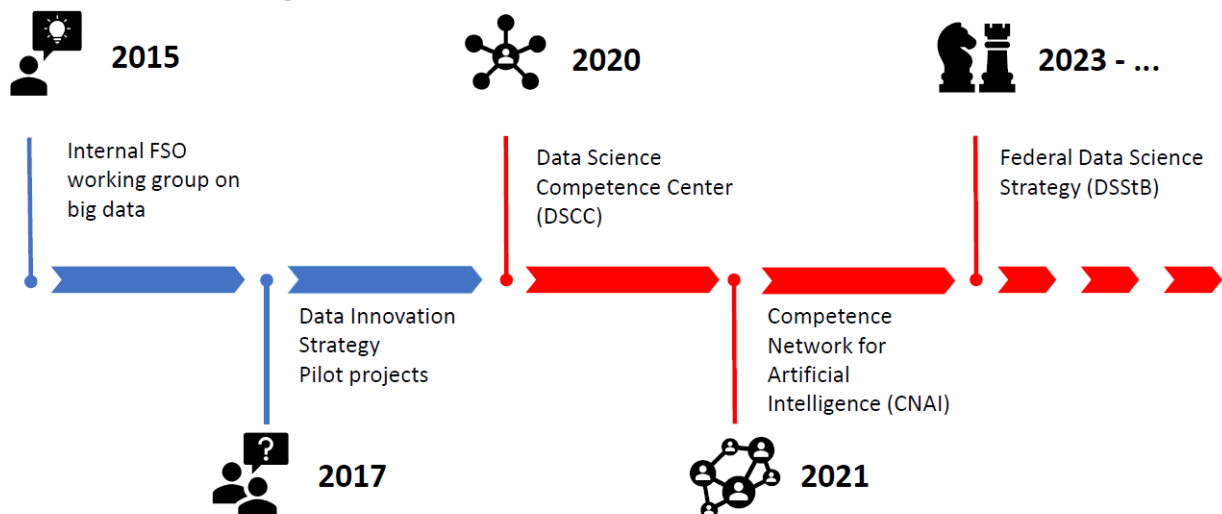
This is the timeline of the creation of our data science competence center; and you may be surprised how we started. We were not really convinced at the beginning what we had to do. We started in 2015; we created a group inside the FSO with methodologists and senior statisticians, and we asked ourselves what could be the impact if we would have to integrate Big Data in our statistical production? We started pilot projects and developed also a data innovation strategy. We wanted to go from a pilot project to experimental statistics and after that go from experimental statistics to official statistics. We abandoned quite a lot of projects because they didn't fulfill the expected quality. I don't want to be too scientific, but we wanted to merge the deductive approach with the inductive approach. We were quite convinced that we needed to move forward with these topics, but we needed people with skills and knowledge that we didn't have inside the FSO.

So, I asked the central government of Switzerland if we could create a data science competence center and the answer was a little bit positive, but we had to cover part of our cost through services to the entire Federal Administration, which means that we would be acting like a start-up inside the FSO. So, we needed to reorganize the FSO. We have since more or less two years three business lines: the

traditional one of official statistics; a new one on data management and data governance; and we have the third one on data science.
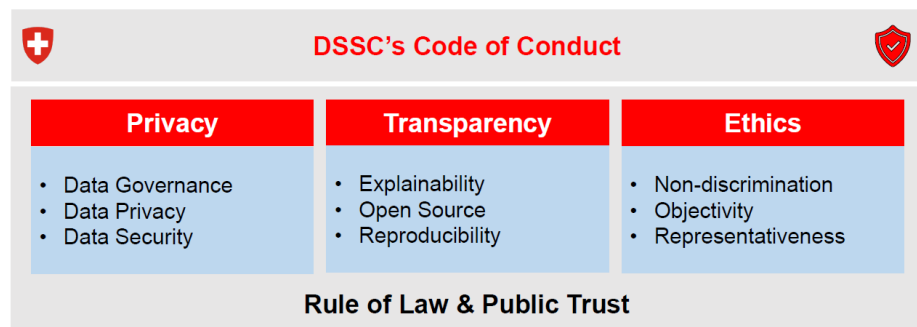
The New FSO since January 1st 2021 - #1
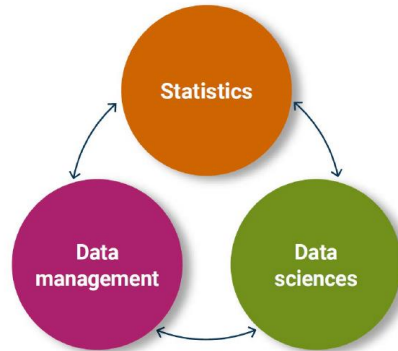
What does it mean that we provide services for all these different topics? We have to deal with different projects, and different type of data, like finance data or mobile phone data. The central government was in favor of creating this data science competence center inside the FSO because official statisticians have a strict code of conduct. In the future this data science center is expected to have many different data sources and more algorithms, and the results of the projects need to respect the code of conduct with strong privacy and ethics rules.

Data Science Under the Rule of Law

Therefore, we have to publish every three months a complete list of projects that we are running inside the federal administration. We have four types of services that we offer, from consultancies up to the training. In the data science competence center we have about 30 staff members, half of them are

DSSC's Code of Conduct		
Privacy	**Transparency**	**Ethics**
• Data Governance • Data Privacy • Data Security	• Explainability • Open Source • Reproducibility	• Non-discrimination • Objectivity • Representativeness
Rule of Law & Public Trust		

methodologists with PhDs in mathematics the other half are data scientists and data engineers, dealing with building the data pipelines.

Provider of DSaaS for the Swiss Public Sector

Data Science as a Service (DSaaS)			
Consulting	**Methodological support**	**Project execution**	**Training**
Advisory and appraisal services on strategic, tactical and operational application of innovative data science methods and procedures.	Methodological support (coaching and on the job training) for data science projects within public administrations.	Complete realization; from problem formulation to Minimal Viable Product of data science projects.	Application-oriented training (off the job training) on data science methods, techniques, practices, technologies and tools.

We also conduct mission-driven research about evidence-based policy making. We have two kinds of projects currently. We have customers who want new insights from data, not especially from Big Data,

and we have customers who want to automatize tasks using AI and machine learning. As examples of such projects, we have managed sensor data from highways. This is real-time data, and we built the data pipelines to deal with these data and show dynamic visualizations of them. We also helped local government administrations with disaggregation of national data. We have also developed in Switzerland some communities of practice on certain topics, and collaborated with universities

Applications of Data Science and AI in the public sector

DSCC's Code of Conduct

Evidence-based policy
- Estimating and assessing the impact of policies
- Improved weather forecasts
- Prediction of the evolution of trends
- Higher resolution statistics
- Real time statistics and monitoring

AI/ML assisted tools
- Anomaly and fraud detection
- Chatbots
- Data matching
- Data preparation
- Data plausibility checks
- Satellite image classification

Safe and scalable computing and storage infrastructures

At the beginning of December 2022, we will publish this data strategy. The idea is to enable the other federal agencies to see the potential of data science and Big Data because currently they don't really see the potential. I will speak very briefly about data governance. Everything started with the "once only" principle for data management within the Swiss government administration. The question that we asked ourselves a few years ago was what should be our role in this new environment. By 2020, we got the mandate from the federal council to create the "National Data Management".

Data Stewardship Model (Governance)

The FSO's Director General is the Swiss Data Steward.
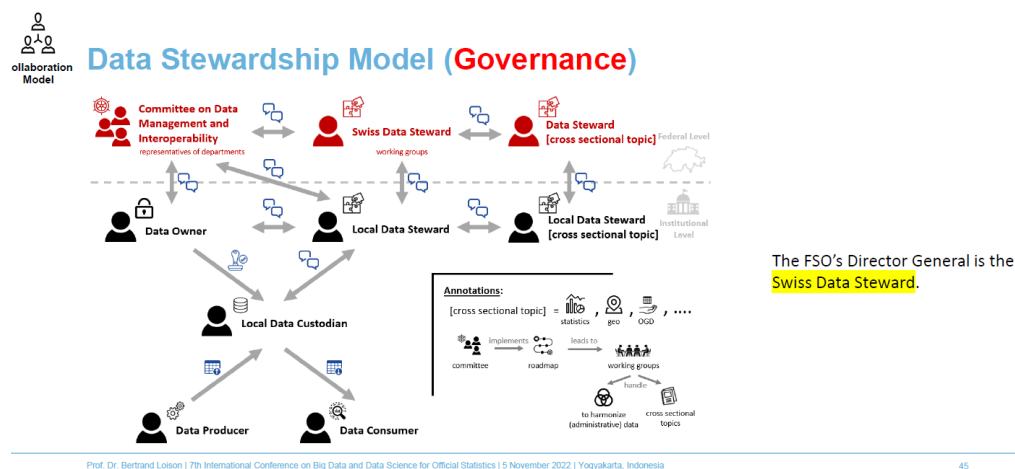
We created a Data Stewardships Model in Switzerland which is quite complicated, but it works, and it is implemented. The FSO Director General is the Swiss Data Steward, and he has a role at the table. He can indicate which data he wants and how we can access the data.

We also created a metadata catalog. I heard today quite a lot of times, that we have to access to the data but that the first thing that we need to know is which kind of data are available. So, therefore we created the metadata platform. Here you can find the interoperability platform: https://www.i14y.admin.ch

Interoperability Platform (Metadata Catalogue)

https://www.i14y.admin.ch
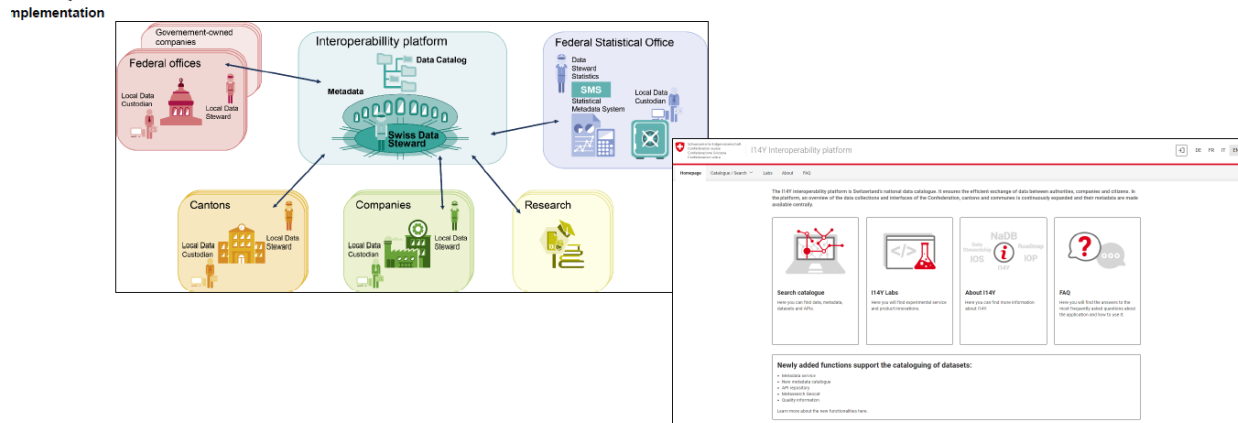
The next issue that we have to solve in Europe and in Switzerland currently is to organize the data in data spaces where, for example, administrative data on debt and other data from the public sector will be organized in accordance with a topic. That's something that we also have to deal with, which, again, is the question what should be the role of official statistician in such a concept of data space.

In Switzerland, we don't have access to the data of the private sector. We can run surveys; we have access to all administrative data inside the Federal Administration, but we don't have a law that gives us access to privately held data. One thing that in my view is important, in this context, is data matching. I'm convinced that in the future matching of units between different data sources can be done in an official way and can be used for official statistics.

Public Trust in the Algorithmic Age
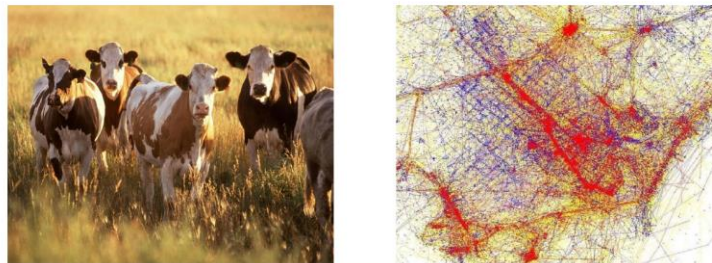
- Fit for the algorithmic age: understanding new issues and learning from more and more data/working out reliable information together.

This will be my conclusion. I'm living in Switzerland where we have always had a lot of cows. We were very good at counting cows. We knew everything about cows. But the times are changing, and we now need to be able to deal with all kinds of new issues, and that's really the reason why we decided to create the data science competence center.

Thank you