Podcast: Global challenges – Global solutions

Transcript for:

Can big data help us understand the future of work?

Interview with Tony Bonen, Acting Executive Director of Canada's Labour Market Information Council, and Matthieu Charpe, Senior Economist at the ILO's Employment Policy Department

Introduction by host:

Hello, and welcome to the ILO Employment Policy Department's podcast series Global challenges - Global solution: The future of work. I'm your host Tom Netter.

Today, we're hearing a lot about something called big data. Can big data help us understand the trends in the world of work? And can it help us figure out the right policies to tackle these issues? To get answers to these questions, let's ask the experts.

My first guest is Tony Bonen, Acting Executive Director of Canada's Labour Market Information Council. We'll look at the challenges and opportunities in leveraging online job ads to track labour shortages.

Then, we'll be talking to Matthieu Charpe, Senior Economist at the ILO's Employment Policy Department on how satellites and the imagery of lights on

earth, can be used to measure economic growth in developed and developing countries.

Tony, Matthieu, welcome to the program.

I'd like to start with Tony. First, many thanks for joining us today. I'm sure some listeners are wondering just what we mean by big data in the context of online job vacancies. Can you give us a brief overview of what it is in this context, how this information is gathered, and how it works in analyzing workplace trends?

Tony: Yeah, for sure. Thanks Tom. So big data can mean many things in different context. Broadly, it refers to large data sets of course but traditionally datasets that are too large to be handled by standard data processing applications. Typically, though, there are three main dimensions when thinking about big data. And that is of volume, so the amount of gigabytes or terabytes or petabytes taken up in the storage; the variety of that data; and then the velocity, the speed at which it comes. So is it real time, near real time, or is it more periodic like more traditional data sources are.

So, when we think about online job posting data, whether that's job posting CVs or other information from the web, there's been a huge explosion of this over the past two decades or so. And during COVID, of course, when many sectors would traditionally rely more on sort of word of mouth for hiring, many of them had to switch to online hiring as well. So, over the past couple of years, we've seen even further growth of this kind of information.

As I mentioned, the information on job postings online is gathered principally through web scraping. So that means actually having algorithms that target different websites, explore what's out there, access that information, that raw text, and then downloads it. That's always the first step.

So that variety of information needs to be organized and it's transformed into organized data we can use through what's called natural language processing algorithms or NLP. So, these are systems that essentially turn that raw text into vectors into numerical structures that allow us to group that information. So, you can group the information, that raw text, into preexisting categories. So, this would be like occupations and so on, or industries associated with job posting. Or,

also very importantly skills or what we always refer to as a work requirement. So, a sort of broader category of skills.

And then the final dimension here is the velocity, which, as you can imagine, is one of the main aspects, the main benefits of online job postings and CV data. That you can collect this data, basically, daily and the frequency of it being updated daily or weekly at the longest time span is a huge improvement on what we can get from traditional data sources.

Thanks, Tony. What are the specific challenges in Canada, and how is big data being used to identify and track them? And how has this situation been affected by the COVID-19 pandemic?

Tony: Yes. So, during COVID-19, we turned to using online job posting data much more frequently than we had before. Prior to the pandemic, we were using this data for the skills-related information. Or as I said before, the work requirement related information. What skills are associated with what jobs, what skills are in demand. Skills data is very difficult to obtain otherwise. You don't observe a skill directly, unlike maybe somebody's wage or employment status.

During the pandemic, though, things were changing so rapidly at that time that the near real time aspect of online job posting data gave us a way to view what was happening even faster than sources like the Labour Force Survey, which in Canada is a very rapid as close to real time as you can get with surveys coming out just a few weeks after the reference period.

Now, we were tracking online job postings through the pandemic to understand the level of demand, the recovery, where the impacts were hitting, according to the different lockdown situations that were being imposed through different waves of the pandemic which was, in our case, very specific to the province in question, since the provinces here in Canada were implementing their own rules and regulations. And so it gave us a way to track that, again, in near real time.

But the online job posting data didn't feed in directly to the sort of analysis of what was going on in the labour market because as there are key limitations that we're still working through. The big ones being, of course, how accurate is the information that's being collected. How reliable are these initial job posting websites that you're extracting from, as well as the processing and tagging of that information to different sources of information, different structures.

And then the big one, the big challenge that we're still working to address is the skewness or representativeness of the data. That there's this skew towards professional knowledge working type of jobs and a lower representation of skilled trades, for example, in these online job posting ads.

How effective are official data sources in identifying unmet demand and available labour supply, and can big data help us get a more accurate picture? I mean, can it tell us about the specific job markets for jobs like teachers, hairstylists and barbers, labourers, and transport workers, and help identify skills needs and skills demanded by prospective employers?

Tony: That is a very critical question. And it's what we're looking at right now. I think there's a lot of opportunity here with online job post data on the skill demand side. There is some opportunity as well on the skill supply side in using CV data or job seeker profiles on certain websites. That's a bit of a trickier one, and we've mostly focused on that demand side. And this is going to be increasingly important as we try and track labour market tightness here in Canada and elsewhere.

There's obviously a lot of talk right now about labour shortages, which is a bit of a problematic term in and of itself, but it does speak to the shift that we're seeing in labour markets from one of relatively easy abundant supply of labour in the past decades to a much tighter labour market and much higher demand relative to the supply of those key skills that are out there.

And so, as I mentioned before, it's really that skill dimension that is unique to online job postings. Now, I think we can make significant inroads in tracking labour market tightness or slack using online job posting data. And you can use this as a more real time source than what's currently available through vacancy surveys. But you do run into this problem again of this representativeness of the data.

Now, what I would say is that, in general, these alternative sources, unofficial sources, are always a compliment to existing official data sources. They shouldn't replace them.

Now the challenge there to your question is the official vacancy survey can only allow you to get so local and so granular as you dig down into that data within a particular city. And say, for a particular occupation, you quickly run out of sample

size and you don't have a significant number of observations to really have a good view on the number of vacancies.

Okay. Well, then looking ahead, I mean, what are some of the policy implications stemming from big data analysis that we can expect in the future of work?

Tony: Yeah, so I think the big one is definitely tracking the skills that are in demand. So, at the Labour Market Information Council, one of the big things we have tried to distinguish is between a labour shortage and a skill shortage. A labour shortage being a lack of bodies coming to the location of work to do the work. A skill shortage being well, you might have applicants and candidates coming through your door or virtually through your website, but none of them have the requisite skills to actually perform the duties that are required.

Now, skill information is difficult to get at and online job postings don't have perfect skill information, but it's much better information than what is otherwise available. In particular, the current set of standard labour market products, so labour force surveys, vacancy surveys, and so on, say nothing about skills that are in demand. And that gap has been traditionally filled with static frameworks, the famous one being Onet from the United States. In Canada, they're developing one here called Oasis that links occupations and those categories to a predefined fixed set of skills.

Now that's very, very useful in a lot of ways, but it is lacking because the variance of the skill set associated with the job doesn't change over time and doesn't change by region.

Okay. So to wrap up, more specifically, can you comment in general on the potential for big data to contribute to key development areas?

Tony: The Labour Market Information Council is a Pan-Canadian organization. So, we don't focus on international dimensions of labour markets, although we certainly work closely with partners at the ILO and elsewhere. But in broad strokes, yes, I do see that there's a big opportunity for online job postings to help bridge the gap between the required labour market information and the need for robust statistical infrastructures that any policy-maker anywhere in the world needs.

Now specifically, I want to say it can help bridge the gap and help avoid the leapfrogging term, which is maybe somewhat similar. The famous example of cell

phones being introduced in many countries before wired telephone systems were implemented. And that leapfrogging is because cell phones effectively replaced landlines; boomers notwithstanding of course. But the same is not true when it comes to online job postings, right?

They are a value added to robust statistical system. So, using them is a very helpful thing, but not a replacement for surveys and other administrative data sources that can help guide labour market information systems, identify training needs. And critically for us, we use a lot of that official labour market data sources as a benchmark to consider how reliable or not online job posting and other big data sources are.

Thanks Tony. Now, I'd like to shift our focus a bit. We've all looked out the window of a plane travelling high in the sky at night and seen vast areas of darkness punctuated by points of light from villages to towns and cities. In a similar way, another element of big data involves using satellite imagery to measure labour market/economic activity from space.

Matthieu, thanks for joining us today. Can you tell us about this application of big data and how it works?

Matthieu: Hello Tom. Nighttime lights is part of remote sensing. So, it needs a category of data that is collected from above the earth's surface by a plane or a satellite. In this specific case, it's the amount of lights at night that can be observed from out of space by your satellite. And the idea here is that you measure a light intensity at night, and you use that measure to be the indirect measure of economic activity. So originally, the satellite was designed to detect a cloud at night. But side effect was that you could observe the lights from human activities.

So why is it, is it big data? It's big data because you have, you can observe every location on the planets every day at a given time to the exclusion of the north and south poles. So you've got a very big geographic coverage. The geographic detail is very small. It's very detailed. So you can observe lights, for every square that is one kilometer by one kilometer. And then you can aggregate this information at the level of analysis of interest, whether it's villages, counties, regions, states. You have data dating back to 1992. You have this information on a daily frequency, but you can then aggregate these daily frequencies into monthly or yearly frequencies. And you have this information in near real time. And lastly, it's also big data because it's free and accessed from your computer.

Thanks, Matthieu, for your interesting response. What are some of the sources of light at night? Cities and towns? Roads? Airports? And how useful are they in forecasting growth, in both developed and developing economies?

Matthieu: I'm going to start by telling you what is the idea behind using nighttime lights to measure economic activities. So, the basic idea is very simple is that higher income are translated into a higher electricity consumption; and therefore, more light intensity at night. Development economists have used for a very long time electricity company records to build a measure of income growth in countries where there's no official statistics. But the advantage of this approach is that with satellite, you have a measure for the entire globe at the high time frequencies and at the very detailed geographic levels, which means it's more information. But the idea is the same.

What are the sources of lights? So, the satellite captures all light at night, which includes light from human activities and non-human activities. As you said in your questions, human lights includes building lights, street lights, roads, and airport, but there's also a lot of sources of non-human light, which include moon lights, boreal activities, or sunlight. And so, it means there's a heavy data cleaning process to differentiate between the two. For instance, when there's a forest fire, it emits lights, and it has to be cleaned from the data.

So how does it work? So, the high-income countries versus low-income countries. So, the idea is to address the problem of underfunded statistical agencies in low-income countries that have difficulties in producing official statistics. So, the application is intended to follow income countries. However, because there is a very fine geographic disaggregation, a lot of work has also been application in high income countries.

So, how does light intensity track income growth and GDP? The first point is that it's used to look at GDP growth and the GDP level because they are cultural habiting electricity consumption. So, what you want to measure is the change over time, not the level.

There is a seminal paper written 10 years ago in 2012. And that showed that the relationship between light intensity and income is very strong and it's been confirmed by many, many subsequent studies. One result is that this relationship is stronger in low-income countries rather than in high income countries. The idea behind that result is that in low-income countries, there is a constraint on the electricity capacity of the country. So as income grows, you have this constraint that is released, and you have a stronger relationship between the two variables.

Matthieu, thanks again. This may seem like a naïve question, but what if it's cloudy or raining? How can this way of measuring economic activity be applied consistently or in the long term? In other words, what are some of the limitations and challenges of this approach?

Matthieu: So, there are factors. So, there's a different between the lights that are emitted on the earth's surface and what the satellite captures. So, the first thing is, what I've mentioned, is the difference between human lights and non-human lights, but there are other things. One factor is cloud because when there are clouds at night, it means the satellite measures less light intensity. So, clouds dampen light intensities. So, cloud is a very, very important factor.

The other thing also is related to the technology in the satellite, is the satellite capable of detecting low light, or is the satellite capable of measuring very bright light in city centers? But coming back to the question of clouds, there are some areas on earth that have very little cloud- free nights. So, what you need to do when you use this type of data is to make sure that the observations that you are using are observations that corresponds to a cloud-free observation. And this comes together with a database. So, you can use what we call a "mask" and exclude some of these non-cloud free nights.

But I would like to say that the main success of this approach is what people call "the power of averaging." So, usually people use annual data at the country level. So, it means that you have taken the average of daily observation at a very small geographic detail, and you have taken the average over time and over the country. And that's what makes this approach interesting is that behind a data at a country level for a given year, you have an enormous amount of information behind it.

Well, this sounds like ground-breaking stuff. Can you give us a few specific examples of how this approach can help shape policy measures? And would this work for other development areas of interest, such as measuring climate change, population growth and urbanization?

Matthieu: This approach is recent, but not new, as it is 10 years old now. It's very popular in development economics and it's increasingly popular in international organizations, such as the IMF or the World Bank. Potentially, you've got a large array of application as it intends to be the proxy for economic growth and it's the main question asked by economists, in general. The original application was to improve GDP statistics. So, there's been a lot of work on this topic. But another

popular approach, for instance, to look at climate change. Questions such as, how does typhoon impact nighttime light and economic activities?

Recently, there's been some work using mobilizing nighttime lights to track the recovery from the pandemic. There are some examples in India and Morocco. And yet, I know I can list some current application. There is, as well, some applications to track the recovery from the pandemic in Asia; as we've seen with the pandemic lockdown measures, they are local measures.

Usually, you do a lockdown of a given regions, a given city, and you try to differentiate between areas that are impacted by the pandemic and other areas. So, it's important to have a disaggregated source of information to measure income growth. And that's where nighttime night has been very useful to track the recovery from the pandemic. Also, one advantage is that because you have daily observations, you can build some high frequency observation of the recovery to the pandemic.

Another application is to use nighttime lights to study the impact of conflict and how they impact jobs in Sub-Saharan Africa. The idea here is to combine nighttime lights with geo-reference database on conflict and to see okay, when there is a conflict, how do nighttime lights react and how can you translate that into employment losses?

The last questions, last area of work in your organization, has been to look at growth convergence in Sub-Saharan Africa. So here, the question is whether poor areas in Africa have been catching up with relatively richer areas or they've been diverging. And, you can do that by mobilizing nighttime light, as you can build a local measure of income growth.

Lastly, I think that will be a very important potential application for the organization would be to use this information to improve how we measure employment growth across countries.

Mathieu and Tony, thanks for these fascinating insights. We'll all be thinking about this next time we check the help-wanted ads or see a bright light flashing across the sky. We've heard some interesting applications of big data that are both celestial as well as very down to earth. And they seem to show that big data can help us improve policy making and service delivery, prioritize scarce resources and make better choices as we address the challenges we'll be facing as we advance into the future of work. I'm Tom Netter and you're listening to the new series of podcasts from the ILO Employment Policy Department on Global challenges, Global solutions: The future of work.