

I have to say it's an honor to be here, this is my first ever conference talk. I did not imagine wearing a microphone and I did not imagine an auditorium, so I'm very, very honored. As you can tell, I'm going to kind of guide you through, sort of giving you an overview of who is a data journalist nowadays, what are the industry characteristics, based on a survey or two that I've carried out. Just briefly about me, I'm a PhD candidate at the Surrey Institute for People-Centered Artificial Intelligence. My focus there is really news deserts, and the research that I'm talking about today instead comes from work I've done while at the European Journalism Center, previously to starting my PhD. Aside from that, I've also worked as a data journalist at the Sheldon Studio, which is an information design studio in Italy, they do really cool projects, so if you haven't heard about them, I recommend checking them out. So without further ado, let's introduce the survey. The survey is called the State of Data Journalism. I don't know if many of you have come across this survey before, it's by [datajournalism.com](http://datajournalism.com), which is a child project of the European Journalism Center, which is a non-profit that aims to deliver funding and training to journalists worldwide. The child project, [datajournalism.com](http://datajournalism.com), is kind of a community gatherer for people interested in data journalism, and two years ago when I joined them, there was this kind of desire to create a survey that would kind of measure what's happening, give us an overview of this growing field. At the time, this was also felt like it was exacerbated by the pandemic, that it put data journalism into the spotlight, and yeah, and also we felt there was generally a lack of global large sample-based survey studies. There are survey studies about data journalism, but perhaps they tend to focus on small regions, and also they tend to stick within academia, so we wanted to do something that we could spread kind of to a global community. I think an advantage for us at data journalism to be able to do that was the fact that [datajournalism.com](http://datajournalism.com) has an amazing network of established names in the community that have contributed onto the podcast that we have, and written long reads for us, or contributed chapters to the handbook of data journalism, and then also that a lot of newcomers to the field, or people interested in the field, perhaps students, use [datajournalism.com](http://datajournalism.com) as a sort of learning platform and a means to kind of know what's happening in the community. And yeah, ultimately a desire to repeat this effort over time, and then be able to capture sort of over time trends. So this is kind of an overview of the survey. The banner at the top should have 2021 and 2022, that's the two editions, and the third one should say comparative elements, just for reference. So you can see the sample is quite large for the two surveys. The numbers in the brackets indicate who we actually ended up using in the analysis, and the number next to it is who took the survey. We ended up discarding some participants due to incomplete submissions, or some people somehow filing the survey twice or three times. We are grateful for their time, but we only wanted one questionnaire. And when it came to the question, it was a really lengthy questionnaire, because when you try to measure the field as a whole, there are so many things to kind of take into account, and these are the sections that we included. In the second edition of the survey, we wanted to add some questions about the Russia-Ukraine coverage. Specifically, it meant, for kind of trying to keep the questionnaire short, that we took away some questions from other sections. So when you look at the two surveys, there are sort of 50 questions that you could compare longitudinally. And when it comes to the people in the survey, we put a question in the second edition of saying, did you take the survey last year? And 411

people replied that they had. When I kind of wanted to match the same person between the two survey editions, I was able to do that myself only for 302 individuals. In terms of impact, the survey has been each edition kind of the subject of a panel discussion at the International Journalism Festival in Perugia, and then a journal article written by myself and a few colleagues came out in February 2023 about the first survey. So this is really the first time that we look at the two in perspective and in relation to each other. So there are kind of two approaches I think we can adopt when it comes to longitudinal analysis. The first one is kind of assuming that the sampling strategy was good and representative enough that you can just compare the results from the first survey to the one of the second survey. And I think it's a good method to capture overall trends. And the second one is kind of what I was mentioning earlier of finding the same individual between the two surveys and say, okay, people replied A in 2021, but then they went on to reply B in 2022 and observe this kind of within individual changes. So I'm going to kind of give you an overview of the overall trends to begin with. The spoiler alert is that things were very consistent across the two editions. For me it was also comforting as a sort of sampling methodology. I feel like we kind of got a good substantial sample, that our answers were very stable. Nonetheless, a few developments were standing out. So starting off with the demographics, the first thing that we noticed across both editions was that the majority of people in the field are male. There were two out of five were women. This was very much the same in the two editions and it also kind of mimics previous findings by none other, my supervisors, Bahari and Haravi did a previous survey and found the same. However, women are younger and they're mostly in education, like a greater share of them is in education than men. So perhaps this is something that is going to change over the years. They're also geographically unevenly distributed. There is no legend there, but I can tell you that the US had the highest share and that's around 11% of all of the respondents. And then we had Italy and the UK following, Germany and Spain, we were from something like between 5% and 9%. And despite this being the fact, actually a little note about that, I feel perhaps many believe that data journalism is concentrated in countries where there are better resources, like organisations with better resources. But to me it was also maybe, I kind of put a question mark there because I think it's just harder to reach populations in areas where maybe data journalism doesn't have such a long tradition or where there are linguistical barriers in terms of reaching them. But nonetheless, we've seen some changes between the two editions. And the main one was that countries in Africa and Asia were rapidly growing at a faster pace than the countries in Europe or in North America, namely India and Nigeria actually entered the top five, they were sharing the fifth place in terms of the number of data journalists that they have, that they work there, whereas in 2021 the top five was just Western countries. So I think that's something to keep an eye out for. When it comes to employment, there's a bit of a gender gap in terms of leadership positions. When we surveyed, when we put out the survey, we asked people to kind of categorise themselves, whether they were full-time employees or if they were part-timers, if they were students, if they were trainers and educators. And when it came to team leads and editors, we saw that there were less women than men overall. And then generally as an industry, most people tend not to have been in it for a long time, over three out of five had under five years of experience. And another remarkable fact I thought was the fact that over one in three, if I remember

correctly, 35% said that they were solely self-taught in terms of data journalism. There were some changes in terms of how people are forming themselves, namely we saw that workplace training is increasing as a means to become formed as a data journalist, but I think it was that begs the question about who then is getting this training, because that's perhaps not freelancers or students. And instead higher education and formal educational courses saw quite a big drop. And there was also a drop in the share of people who are self-employed, a difference of five percentage points between the two survey editions, meaning that perhaps people went from, that there are now perhaps less freelancers. And instead at skills and tools, one pattern that was very obvious between the two survey editions is that people deemed themselves much more comfortable at journalism than any other data skill. And I think this ties very well with another finding that we saw, which is that more than half wishes to be upskilled in data journalism and data related tasks. And in fact, less than half said that they were upskilled in these tasks, while the majority was being upskilled in journalism. So I think this is an indication that people are demanding and wanting training in things like data visualization, data analysis, statistics, and also machine learning. And another curious fact I think was seeing that only one in four uses programming. I think a lot of, or I think there is this kind of pressure among data journalists to kind of be very techie, but when you see how people do their job, really it's the minority that codes. Looking instead at work practices, I think one, this was a point that was raised as well at the Journalism Festival. We saw that like 40% said they'd cover four beats or more, and I think that's quite remarkable because if you think about the amount of domain expertise that you want to have as a reporter, having to cover so many different beats can be quite the pressure. And then maybe this is perhaps less surprising, but we found that 50% of individuals cover local or, sorry, international or national stories, while only one in four covers local. I think I'm already running out of time. Is five minutes to the end end-end, or is five minutes to questions? Five minutes to questions. Ah, okay, great. Okay, then I'm maybe not so, all right, all right. Okay, so in terms of the challenges, I think the one thing that really we should take away is the fact that there are huge regional inequalities in terms of data access, access to quality data. You see this in the plot here. We asked about how they would rank access to local national data and the quality. Essentially, if you see a country in this lower quadrant, it's going to be in that one as well. The correlations were very stable. And access to quality data was also kind of indicated as the top hurdle. And if you think about the fact that perhaps data is the one ingredient that we need in some step of making a story, that's quite important. Talking very briefly about the pandemic, I think one of the most interesting insights was the fact that one in four growing to 40% in the second edition said that they became involved in data journalism as a result of the pandemic. And I think this was a shortcoming on our end of not kind of digging deeper into the why. Was this because people saw a need for data journalists and they decided to shift themselves to data journalism from a standard reporter position or was this an organizational pressure from editors, newsrooms? Was this a strategic demand for more data journalists around? And then when it comes to kind of like the personal effects of the pandemic, a lot of people felt negatively affected. The largest group was actually not deeming themselves negatively affected when it came to work pressure, time pressure, resources, and workload. When you combine the people who were somewhat negatively affected or very negatively affected, that

group was the largest overall. But we also see that it's kind of like shifting over time. The blue curve is 2022, and you can see it kind of shifting more towards neutral or positive. Just very quickly, let's now look at these 300 individuals that I was able to match between the two surveys. I put some demographics there to show that these 300 individuals are quite representatives of the overall sample. And the first thing that I noticed was that a lot of people changed jobs, like there was like 37%. I don't know if that's like high or low to me, it seemed high. And the group that was most likely to change job were editors. In fact, 60% of editors that took the survey in the two editions changed to something else. And they mostly went to become full-time employees or perhaps something else. And then we also saw that of those who became editors, most were full-time employees. Perhaps that's not surprising if you think about promotions and so. And then when it comes instead to the other highest paying job, which is being an educator in the industry, as we saw through an income question that we put, we see that these come from a variety of backgrounds. So there is less of like a determined path for these people. And lastly, we saw that also freelancers tend to essentially just switch from part-time to full-time or vice versa, without changing kind of occupation drastically as others would do. Okay, okay. We asked people if they felt they were going to be wanting to be like in the industry in five years' time. And while the majority, I think it was like around 80% or 75%, said they wanted to stay in the industry in five years' time in both editions, we saw that like a much larger group who said they wanted to stay in the second edition of the survey, they actually told us they wanted to leave at a much higher rate than the people who first said no, then decided they would have liked to stay. And I can't really, I don't really know why that is the case, but I think that's kind of interesting. It's one in 10 that now thinks about leaving the industry. We saw some organizational shifts as well, with people going from either very large companies to kind of more medium-sized one, or from the very smallest companies to the medium-sized ones. And then lastly, we saw some changes in terms of work mode or pandemic long-term effects. The first finding that we saw is that basically the share of people that work from home is kind of like the same as people who work hybrid. But I guess a reassuring pattern that we saw is that more people are re-entering the office now than vice versa. My text has become so tiny, I don't know if you guys can read it, but I essentially put out there a couple of considerations when you carry out a study like this. Well, first, I would like to say that there is so much data, and I would have even liked to ask so many more questions, but you know, you can only ask people for so much of their time for free. But there is so much data that can be investigated further. In one way, you can, you know, make a, you can just get the subset of people from the country you're interested in, or the region that you're interested in, and study that. I think it's, yeah, one inherent challenge of the method, all in all, I think, is being able to capture a representative sample of data journalists. This is not information that is kind of publicly known, and yeah, we kind of had the aim of, you know, leveraging this big network that we have, and kind of hope that the sample would be representative. But we're limited in how much we can verify that. And then the last thing I would say is that the service as a method themselves are not really allowed to answer why, to kind of explain why we see what we see, unless we kind of add many more questions. So I wish I had more explanation as to why we see the data points that we have, but I think that kind of is an inspiration to dig further, ourselves and other researchers out there. That's it. If you

want to kind of explore the data, it's all interactive on the [digitaljournalist.com](https://digitaljournalist.com) website. You can also grab the data yourself on GitHub, or if you want to, you can read the paper that we've wrote about it. Thank you so much. And we have just a minute for questions. Matias. Thank you, Simona, it was a really great presentation. I know your work, and it's amazing. So a quick question that always comes to my mind is about Italy. And I think there is somehow this number of answers that you have that is way bigger than other countries that I think data journalism is strong, it's because of your network. So do you think there is a way of European journalism center, or you work with some people that have a strong network like you in other countries to make these numbers more, for example, he's a really big name in Brazil. So for example, to have more answers that we could be more representative in this data. Yeah, I think, I mean, without a doubt, I think better research is possible to give further attention to some regions, partnership could be created, leveraging local networks that could also help us translate the survey. Our survey was available in four languages, and that for us was a way to promote it to some region. Italian, Spanish and Arabic. But nonetheless, I think it can go further. But it's also a matter of resources as well, at the end of the day. We can discuss later. Yeah. Otherwise, I'm ready to. Yes. All right. All right. Thank you so much. Thank you.