

Module 1 Video 2: The Seven Steps in the Data Equity Framework

[00:00:01] Hi, welcome back to video number two in our module, where we're talking about the seven step data equity framework and how you can harness the power of this framework to improve the equity and ethics in your data journalism work. So there are seven steps in this data's lifecycle and that starts from funding and goes all the way through communication and distribution.

[00:00:28] So the first step is funding. So if you are using data or any sort in your data journalism product or story, you need to know who paid for the collection of that data. For example, you can find really good research coming out of Goldman Sachs for how well they helped vulnerable young people get into college and stay into college. They have research on that. You will not find research that shows how many students were forced to drop out of college or forgo it altogether due to the 2007, 2008 economic crisis brought about in part by Goldman Sachs.

[00:01:15] Why do some of that data exist and some of that data not exist? That's because of funding. Goldman Sachs paid for one of those research studies, one of those data sets to exist and did not fund one of those other research studies. Funding has a direct effect on what data exists and from the data that exists, what it actually says. So if you're going to share data or statistics or research in your data journalism article, it's very important that, you know. Who funded? Who paid for that data to exist? Because it's quite expensive to get data to exist. And somehow your audience needs to be able to find that information, whether it's through a footnote or a link. You need to make transparent. Who paid for this data that you're citing?

[00:02:08] OK. Number two is motivation. Motivation is why does the data that you're citing exist? Who decided that this data would exist and what purpose is it serving?

[00:02:25] Step three is project design. So the data that exists was collected by somebody who designed the way that it would be collected. So, for example, all the funding and all the motivation that led to the data being created. Somebody had to operationalize those ideas. Some people had to decide why and how that data set would come into existence. For example, if your data project is on poverty reduction. Who has designed a data that studies why poor people make certain decisions with their money and who has designed data and research that examines how rich people spend their money. This is called studying up. So studying up is a very important concept in terms of project design. So if you're a data journalist and you're going to use data, you need to find out before you use the data and document who and how the project design was implemented in the data project that you are citing. Another really, really good example of data and project design is who got to define the sampling area. Now, we're going to talk a little bit more later about what sampling is. But essentially, we want to know what was the sample? Who got to design? How the data was selected? So, for example, if you live and work in North America and you are using polling data or survey data, that data was sampled probably using a colonial project design. So it was sampled along the lines of states or districts or wards or provinces. This map that I'm showing you here on the screen is developed and maintained by a fantastic organization called nativelands.ca. And this shows the boundaries that existed in North America before colonial treaties.

[00:05:06] What if you were able to access or use data that was collected through a project design that was not colonial in nature. That's what something like nativeland.sa allows you to do. So when you're doing data journalism, you have to think about who designed the project and how the project design is influencing the results or the bias that is embedded into the data that you are using and presenting. OK. The next example in terms of project design is a report, a research report that turned into a data story that showed that risky behaviors were being reduced in certain communities and the risky behaviors included smoking, drug abuse and premarital sex. And in this case, the media outlet had quite a problem because lumping smoking, drug abuse and premarital sex into a category called risky behavior is a very value based opinion. Whether or not premarital sex is the same as drug abuse is the same as smoking, is somebody's opinion. And so using data that came from this study in your data journalism story, embeds the project designs opinion of what is a risky behavior into your data story. So this is another example of how project design as step three is an important step to consider in your data journalism story.

[00:06:57] Phase four is data collection and sourcing. So if you're a data journalist, you know that finding the data that you need is incredibly hard. So you're going to have to get it from somewhere and you're going to have to decide whether or not you're going to trust it.

[00:07:16] Step five is analysis and sometimes as a data journalist you're doing your own analysis and sometimes as the day journalist, you're going to rely on other people's analysis. And analysis itself is not objective. So statistical methods all have a set of assumptions baked into them. If you took like a stats 101 class as an undergrad, you might have heard your teacher talk about the assumptions that are based in each statistical methodology, whether you're going to use a regression or an ANOVA or multi-level model. You don't need to know what any of those things mean. All you need to know is that each one of those is based on a set of assumptions and those assumptions are world views. So we're gonna get into the nitty gritty of analysis in the third module. But it's important, as a data journalist for you to understand that statistical and mathematical analysis are not objective. If you're still confused about that. Go back to the introduction module and watch the video about the average classroom size. And that's a really fundamental example of how even the most basic mathematical analysis has somebody's world view baked into it.

[00:08:34] This does not mean that I'm saying that there's no such thing as truth or that there's no such thing as a fact. What this means is that data science is a science. Data journalism is a good idea. It can give repeatable and accurate answers to your questions without just relying on people's intuitions or emotions to get those answers.

[00:08:59] But that does not mean that data or data science is an objective result or that math doesn't lie. We do embed world views all along the way. So there's a big difference in saying there's no such thing as a fact vs. world views are embedded into data. And we as journalists need to understand how world views are embedded into data and make that transparent in our data journalism stories. Again, go back to the average classroom size video if you are wondering whether math really is objective or not.

[00:09:42] So these are the seven steps in the data journalism cycle. We've done funding, motivation, project design, data collection and analysis. Then once you have your analysis, this is where you, as a data journalist, probably do most of your work in step six and seven, which is interpretation and communication. That's primarily your job as a data journalist. The interpretation and communication.

[00:10:12] Interpretation is a really important step in data journalism and is probably the step where the most equity and ethical issues come up as a journalist, because interpretation is where you need to build a story or embed a narrative into a mathematical or statistical results. Probably you will have heard about the controversy around the COMPAS Algorithm and the COMPAS Algorithm is used or was used in some jurisdictions of talking about whether or not somebody is likely to reoffend.

[00:11:00] ProPublica did a fantastic full length story. Head on over to ProPublica and read the whole thing. However from an interpretation perspective, it's a very simple issue. The interpret, there's a COMPAS Algorithm and the COMPAS Algorithm gives you a score. For sure it gives you a score. And many people. Put the interpretation around the COMPAS score to say that COMPAS gives you a score that predicts how likely it is that a person will reoffend. So that interpretation is extremely problematic, and you can see from just the two quick numbers that are shown here from the ProPublica work that the color of your skin has a lot to do with the score that the COMPAS gives you.

[00:11:58] The interpretation of the COMPAS score is extra is extremely problematic. So there's no doubt that the COMPAS gives you a score. One of the main problems is around the interpretation of the COMPAS score. So at first it was interpreted by some people to believe that COMPAS gives you a score that predicts how likely it is that this person will reoffend. So that's gonna help you decide should this person get bail, should this person get parole, etc. However, what COMPAS really does is gives you a score that predicts how likely it is that this person might be in contact with the police, again, be arrested by those police and not have the money for immediate release.

[00:12:40] You can see how that's an extraordinarily different interpretation of the same number. And it makes this interpretation is much more transparent about what's actually going on in the COMPAS Algorithm. And it also is much more transparent about all the baked in racism, sexism and other problems that might exist in this score. So as a data journalist, it's really important that you ensure that you are not just kind of glossing over an interpretation of a number. But that you're really paying attention to specifically what is the correct interpretation of this number? How mindful, how intentional can you be about the interpretation of whatever data you're going to embed in your story? Because often it's not obvious what an interpretation is. This interpretation is when you take the results, which is the end of step five, the analysis step and you apply meaning. And you embed a narrative around the results that explain them. This is really your job, as a data journalist, and it's where you need to make sure that you are doing your job correctly.

[00:13:59] Then the last part of your job as a data journalist is step seven, communication and distribution. And you might hear a lot about a set of dataviz best practices. And there's a lot of really great research and a lot of support. That's excellent. About what? How humans perceive information, how humans like to read numbers and interpret pictures to help you accurately create dataviz and there's lots and lots of great courses and books about how to create good dataviz. From an ethics and equity point of view it's really important for you to understand that all of these dataviz, best practices are not culturally universal. The vast majority of them have been developed through research that has relied largely on only people that live in Canada, the United States and Western Europe. And lots and lots of those people are in University courses, so what's called court of dataviz best practices are often skewed towards being best practices for people who live in a certain part of the world and am educated to thinking to see in a certain way. I do a lot of my work outside of Canada, the US and Western Europe, and I've learned the hard way many times that the best ways to portray or visualize data in other parts of the world are not the same as they are on Twitter or on the really Western European North American Dataviz Best Practices.

[00:15:47] So we need to be really careful about whose preferences and worldviews were embedding into our dataviz. And we're going to talk a whole module about that four module four.

[00:16:00] So just to review one more time. These are the seven steps: funding, motivation, project design, data collection, analysis, interpretation and communication. You need to think about each one of these steps as you're writing a data journalism story and embedding ethics and equity into that data journalism story. So we'll start looking at some practical examples, doing some readings and talking about some of these specific steps that are most important to data journalism in the following modules.