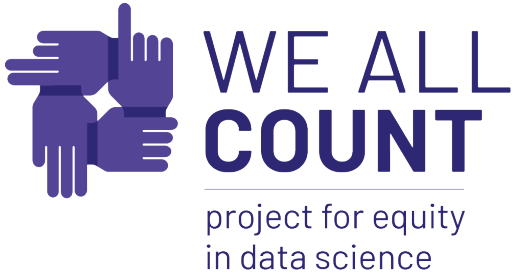# The First 99 Data Equity Choices

WE ALL COUNT
project for equity
in data science

There are many, many choices to be made in any data process, but here's a list of fundamental, nearly universal choices that we think lead to some of the biggest equity imacts.

Remember, level 1: intentionality and level 2 transparency are crucial stepping stones to level 3. We can't align our decisions with our equity priorities if we can't identify and talk about our choices.

Did we make this choice...

Funding		CHOICES	Lvl 1: Intentionally?	Lvl 2: Transparently?	Lvl 3: In alignment with our equity goals?
	1	Who are the people and groups involved in our project?	☐	☐	☐
	2	Who else could we get involved?	☐	☐	☐
	3	Why are we working on this project, not something else? Who chooses?	☐	☐	☐
	4	Who will we give decision making power?	☐	☐	☐
	5	Who will we give meaningful input into decisions?	☐	☐	☐
	6	Who will we say owns the data?	☐	☐	☐
	7	Who will we say owns the data products?	☐	☐	☐
	8	What is the power structure of our project?	☐	☐	☐
	9	What is the data-sharing structure of our project?	☐	☐	☐
	10	Where and how does money flow in our project?	☐	☐	☐
	11	What is the value of the data in our project? To whom?	☐	☐	☐

	12	What is the price of the data in our project? Who sets it?	☐	☐	☐
	13	What should ownership look like?	☐	☐	☐
	14	Who will profit from the data?	☐	☐	☐
Motivation	15	Why are we doing this project?	☐	☐	☐
	16	Who are we trying to prioritize?	☐	☐	☐
	17	What is our specific motivation?	☐	☐	☐
	18	What are the definitions of key terms in our motivation?	☐	☐	☐
	19	What are our non-negotiable goals?	☐	☐	☐
	20	What are our nice-to-have goals?	☐	☐	☐
	21	How much time do we have?	☐	☐	☐
	22	How much money do we have?	☐	☐	☐
	23	How many resources do we have?	☐	☐	☐
	24	How much expertise do we have?	☐	☐	☐
	25	What additional restrictions do we have in our project?	☐	☐	☐
	26	Should we attempt to change to rewards/restrictions of our project?	☐	☐	☐
	27	Who gets to set these parameters?	☐	☐	☐
	28	Who should set our goals and triage them?	☐	☐	☐
	29	How many different directions for this project have we considered?	☐	☐	☐

Project Design				
30	Why are we inclined towards this particular project design?	☐	☐	☐
31	Does our project design prioritize us, our equity priorities, or someone else?	☐	☐	☐
32	What should we count as evidence?	☐	☐	☐
33	Where should we look for data? According to who?	☐	☐	☐
34	What data should we look for? According to who?	☐	☐	☐
35	Who picks what questions are relevant?	☐	☐	☐
36	What hunches/hypotheses are driving our project design? Whose are they?	☐	☐	☐
37	Who is crafting research questions?	☐	☐	☐
38	Where does our research question place expectation, onus or assumptions?	☐	☐	☐
39	How is our definition of success embedded into our research question?	☐	☐	☐
40	What type of question(s) are we setting out to answer?	☐	☐	☐
41	What methodology is the best aligned with our questions and our priorities?	☐	☐	☐
42	Who gets to pick the methodology?	☐	☐	☐
43	Why are we picking one methodology over another?	☐	☐	☐
44	Who will we collect our data from?	☐	☐	☐
45	Who should collect the data?	☐	☐	☐
46	What identity-based data are we interested in and why?	☐	☐	☐
47	What are the consequences of the categories we create/accept?	☐	☐	☐

Data Collection & Sourcing	48	What proxies will we use?	☐	☐	☐
	49	What proxies must we avoid?	☐	☐	☐
	50	How do we want to collect identity data?	☐	☐	☐
	51	How do we plan to use/analyse identity data?	☐	☐	☐
	52	What is our sweetspot for ease, certainty, specificness, and representativeness around identity data?	☐	☐	☐
	53	How will we sample?	☐	☐	☐
	54	What will our sample sizes be?	☐	☐	☐
	55	What is our plan for small sample sizes?	☐	☐	☐
	56	Do we know enough context about the data we are sourcing to use it?	☐	☐	☐
	57	What are the assumptions baked into the dataset we are sourcing?	☐	☐	☐
	58	What are the definitions embedded into the dataset we are sourcing?	☐	☐	☐
	59	What were the parameters (rewards & restrictions) of a dataset we are sourcing?	☐	☐	☐
	60	What was the project structure/power dynamics of a data set we are sourcing?	☐	☐	☐
	61	What type of analysis do we need? (descriptive, causal, predictive?)	☐	☐	☐
	62	What will we choose as our key denominators?	☐	☐	☐
	63	How will we we treat identity categories?	☐	☐	☐
	64	How will we weight averages?	☐	☐	☐
	65	What are the consequences of our resolution of analysis?	☐	☐	☐

Analysis	66	Who should our predictive algorithm be optimized for?	☐	☐	☐
	67	What measure of predictive fairness should we adopt?	☐	☐	☐
	68	What should be our benchmarks for fairness in a model?	☐	☐	☐
	69	What conceptual model (worldview) will we base our mathematical model on?	☐	☐	☐
	70	What variables should being included?	☐	☐	☐
	71	How should we include those variables in our model?	☐	☐	☐
	72	How will we control variables?	☐	☐	☐
	73	What should be our threshold/target for level of certainty?	☐	☐	☐
	74	Who should be making analysis decisions?	☐	☐	☐
	75	Is the way we analysed our data congruent with the goals and choices made in the rest of our process?	☐	☐	☐
Interpretation	76	Can we seperate our results from our interpretation?	☐	☐	☐
	77	How should we reach our initial interpretation of the results?	☐	☐	☐
	78	Does our interpretation maintain our original intended definitions of key terms and catagories?	☐	☐	☐
	79	Is our interpretation supported by the methodology chosen?	☐	☐	☐
	80	Is our interpretation suported by the data we ended up with?	☐	☐	☐
	81	Is our interpretation supported by the type of question we asked?	☐	☐	☐
	82	Is our interpretation a valid output of the analysis we conducted?	☐	☐	☐
	83	Does our interpretation apply to the context we are framing it in?	☐	☐	☐

#	Question			
84	What level of certainty should we attach to our interpretation?	☐	☐	☐
85	Whose point of view should be centered in our narrative?	☐	☐	☐
86	Who will decide on the key takeaway of our narrative?	☐	☐	☐
87	What context should we use for our narrative?	☐	☐	☐
88	Where should we place the onus to change or suggested cause in our narrative?	☐	☐	☐
89	What values should we apply to our narrative?	☐	☐	☐
Communication & Distribution 90	Who is it most important to communicate this information to?	☐	☐	☐
91	Who else might we want to communicate this to?	☐	☐	☐
92	How should we design the content to work best for our priority audience(s)?	☐	☐	☐
93	Who should we optimize our data visualization for?	☐	☐	☐
94	Whose prefered icons, symbols, and aesthetic should we use?	☐	☐	☐
95	Whose prefered language, narrative, and structure should we use?	☐	☐	☐
96	What medium should we use to distribute it?	☐	☐	☐
97	Who are we priviledging with access to the information?	☐	☐	☐
98	Who decides on design, communication, and distribution decisions?	☐	☐	☐
99	What avenues should we create for responses to our work?	☐	☐	☐