Usually this is done by merging several datasets. Many DHS surveys (all?) ask where the household gets their water (piped, well, etc). Then, people often bring in external data from some country or region on expansion of water policies/investments, and use that information to instrument for water access by the households in the DHS data. The water availability roll-out data usually has to come from just shoe leather work: google, talks with government people, reading about programs in various countries, etc. But that information (plausibly exogenous roll out of water access) is just something you have to find on your own (otherwise someone probably would have already done it!).

So the basic idea is something like this, for estimating the causal effects of water access on diarrhea (or whatever outcome): external data for an instrument, water access data from the DHS is the first-stage dependent variable, and then diarrhea (or whatever) would be the final outcome in the second-stage.

That help?