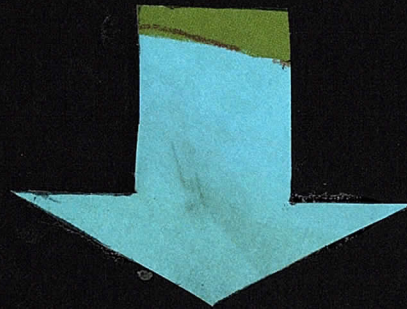What data am I using?

Whose labor produced it?

What biases and assumptions are built into it?

Why am I choosing this particular phenomena for digitization/transcription?

What does the data leave out?

Research:

Is this the soundest data I have access to?