ehRAF quirks:

* Sometimes a passage will appear as a subset of another passage
  + ex:, in the serb culture, in “Healing with mother metaphors: Serbian conjurers' word” the 1989 passage starting with “. Similar round ritual cakes, adorned with…” is a subset of another passage found in the same document
* Sometimes the entire passage will appear twice and will be coded differently between the passages despite the two being the same
  + Ex: in the Kuna culture, in “Kuna ways of speaking: an ethnographic perspective” the 1983 passage of “When you clear a banana farm,’ he says.” Has the OCM coding of [5310, 759, 902] while another of the same passage in the same document has the OCM’s [537, 756, 759, 902]
* A small (0.25%) number of passages are completely blank yet have OCM codes attributed to them. Some are a string of blank passages with the last passage having the word “notes” with the same OCM’s attributed to each blank
  + Ex: in the Klamath culture, in “The Klamath Indians of southwestern Oregon” many passages are blank
  + Ex: in the Azande culture, in “Witchcraft (Mangu) among the A-Zande” the 1929 passage contains no text but has OCM’s
  + Currently, we are removing these in the stats/cleaner
* Some passages have OCM’s lower than 100
  + Ex: in the Saami culture, in “Lapp life and customs: a survey” the 1962 passage “The Lapps have always lived in close contact with nature” has the OCM “77” despite this OCM not being listed as anything even on the ehraf website and may not exist.
* Some passages have OCM’s which do not exist (at least I could not find)
  + Ex: in the Korea culture in “Culture summary: Korea” the 1998 passage “Shamanism is performed by shamans, most of whom are women…” has the OCM 1787 but no further reference to what this means
  + This is not the same as 5310 and 5311 which are subparts of 531 and therefore actual exist.
* When searching for higher order OCM’s like 750, you get every OCM below it (751 through 759) This is not a flaw or mistake and was intended. However, the passages obtained do not always follow systematic conventions.
  + Some passages will ONLY contain the higher order OCM (750)
  + Some passages will contain both the higher and lower order OCMs (750, 751, 753)
  + Some passages will ONLY contain the lower order OCMs (751, 753)
* A number of passages were likely uploaded before larger UTF character sets were implemented meaning that some accented and non-roman characters like potentially âøç would be reduced to roman non-accent variants (aoc) with the text “[unknown]” preceding the character.
  + For example “of the ‘Umar Ma[unknown]hm[unknown]ud” the “[unknown]” is referencing the “h” and “u” having accents to them
  + However, not ALL accented characters removed and this may differ between passages