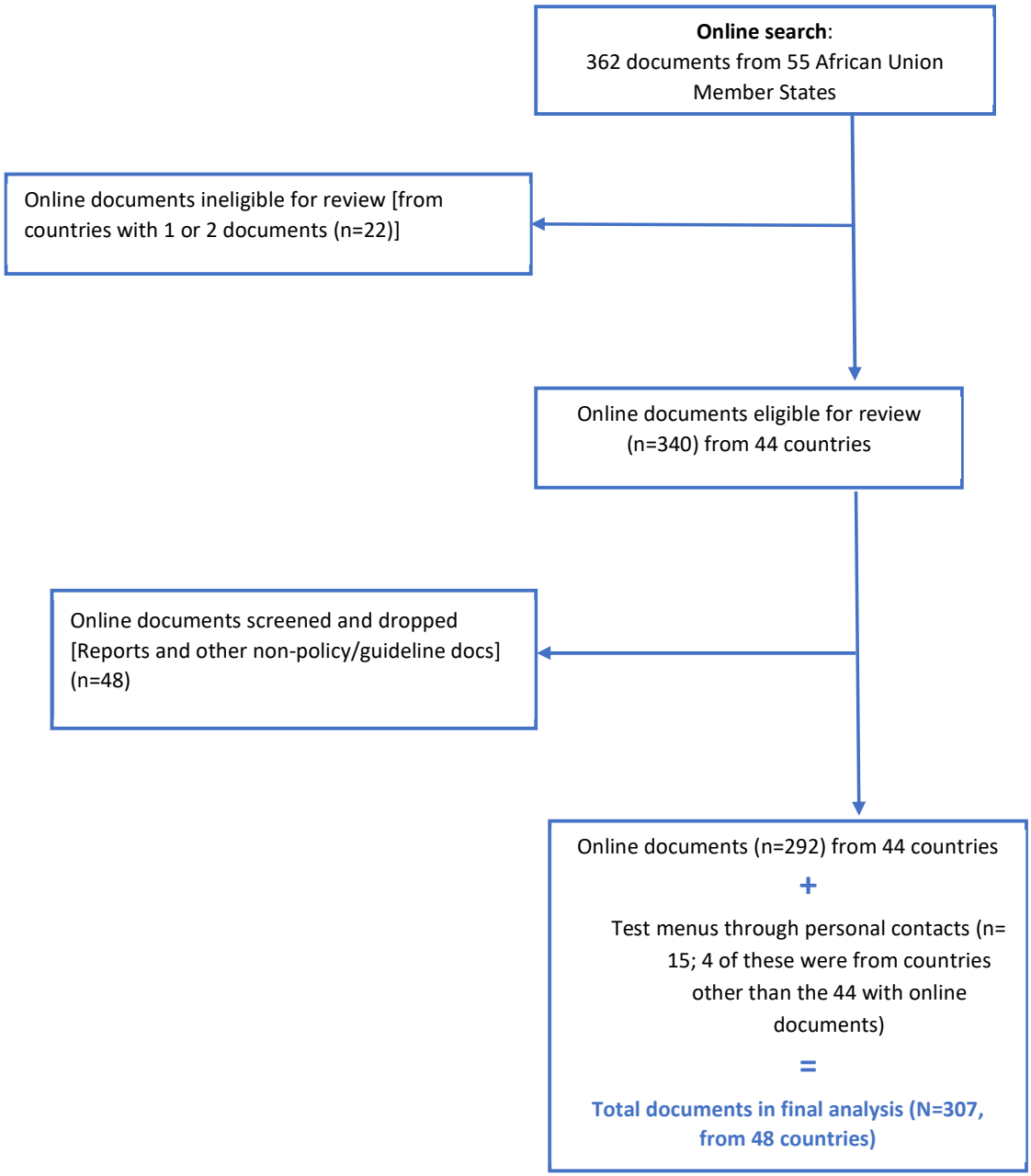S7 Figure. Flow chart sampling of documents