

PubMed Central,
n = 745

PubMed
n = 2,176

Empty abstracts, *n* = 542

Duplicates, *n* = 629

Exclude journals, *n* = 277

Big Data Corpus
n = 1,473

Match PubMed search
n = 14,146

Empty abstracts, *n* = 1541

Big Data duplicates, *n* = 346

Big Data paper, *n* = 14

Big Data Corpus
n = 1,487

non-Big Data Corpus
n = 12,245

