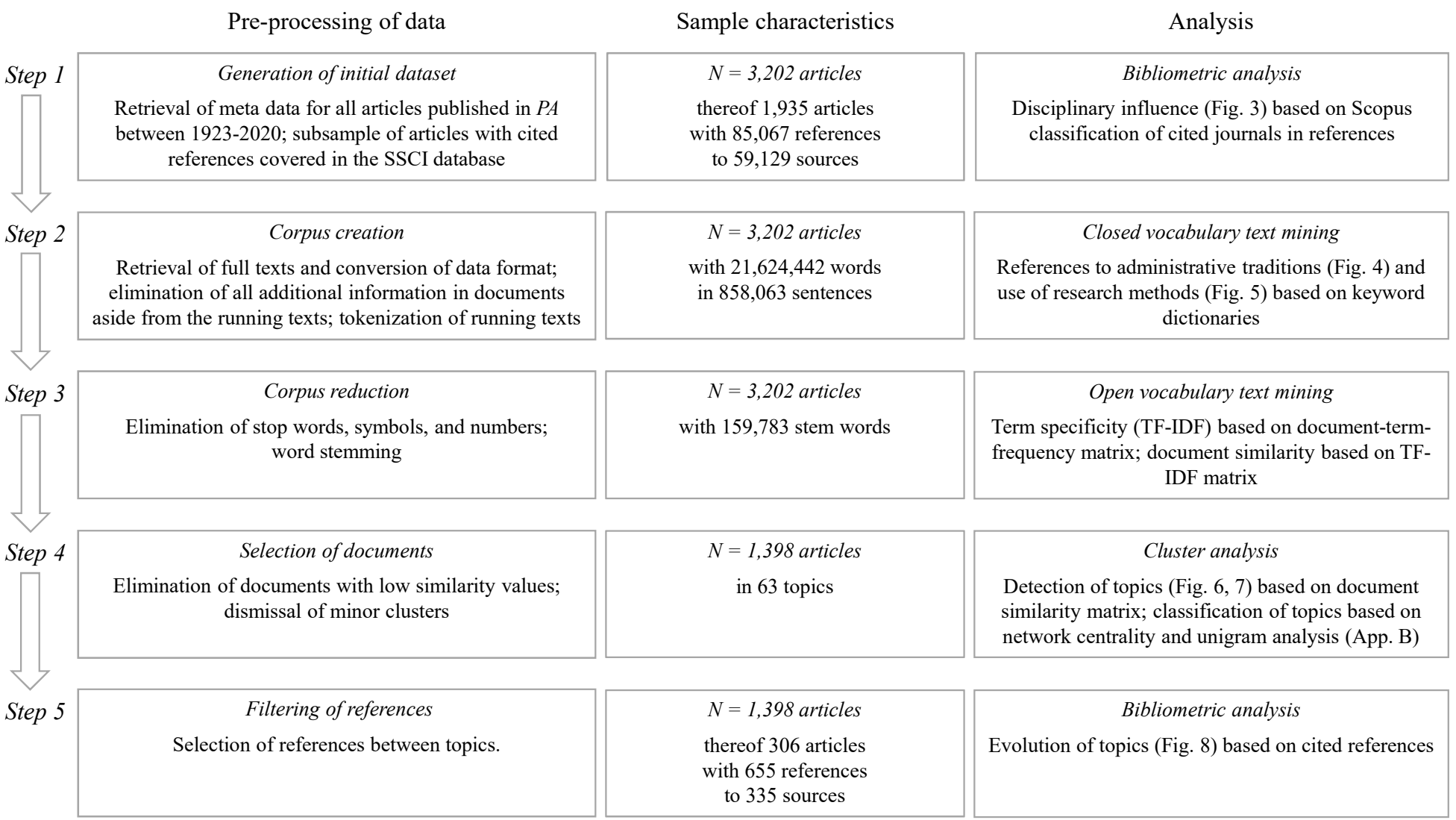
Pre-processing of data
Sample characteristics
Analysis
Step 1
Generation of initial dataset
Retrieval of meta data for all articles published in PA between 1923-2020; subsample of articles with cited references covered in the SSCI database
N = 3,202 articles
thereof 1,935 articles with 85,067 references to 59,129 sources
Bibliometric analysis
Disciplinary influence (Fig. 3) based on Scopus classification of cited journals in references
Step 2
Corpus creation
Retrieval of full texts and conversion of data format; elimination of all additional information in documents aside from the running texts; tokenization of running texts
N = 3,202 articles
with 21,624,442 words in 858,063 sentences
Closed vocabulary text mining
References to administrative traditions (Fig. 4) and use of research methods (Fig. 5) based on keyword dictionaries
Step 3
Corpus reduction
Elimination of stop words, symbols, and numbers; word stemming
N = 3,202 articles
with 159,783 stem words
Open vocabulary text mining
Term specificity (TF-IDF) based on document-term-frequency matrix; document similarity based on TF-IDF matrix
Step 4
Selection of documents
Elimination of documents with low similarity values; dismissal of minor clusters
N = 1,398 articles
in 63 topics
Cluster analysis
Detection of topics (Fig. 6, 7) based on document similarity matrix; classification of topics based on network centrality and unigram analysis (App. B)
Step 5
Filtering of references
Selection of references between topics.
N = 1,398 articles
thereof 306 articles with 655 references to 335 sources
Bibliometric analysis
Evolution of topics (Fig. 8) based on cited references