UNIVERSITY OF MACEDONIA

DEPARTMENT OF APPLIED INFORMATICS

GRADUATE PROGRAM

Bibliometric Analysis on Business Processes

M.Sc. Thesis

of

Viktoria Paschalidou

Thessaloniki, June 2022

Bibliometric Analysis on Business Processes

Viktoria Paschalidou

Diploma in Production and Management Engineering,

School of Engineering, Department of Production and Management Engineering,

Democritus University of Thrace, 2018

M.Sc. Thesis

submitted as a partial fulfilment of the requirements for

THE DEGREE OF MASTER OF SCIENCE IN APPLIED INFORMATICS

Supervisor: Dr Konstantinos Vergidis

Approved by examining board on 28 June 2022

Dr Konstantinos Vergidis	Prof Maro Vlachopoulou	Prof Emmanouil Stiakakis
_____	_____	_____

Abstract

A great amount of research has been already conducted in the business process domain. To investigate the current landscape and the potential implications of the pandemic the aim of the thesis is to examine the degree to which the business process field is shaped by diverse contributions and to discover the trends in the field. The Bibliometric Analysis review method is deployed and the main procedure is complemented by a proposed preprocessing procedure, which supports the profound understanding of all the elements of the analysis. The Co-authorship and Co-occurrence analysis techniques that were conducted reveal the lack of multidimensional contributions in the research field and the low number of publications that address current advances in industry and technology.

Keywords

Business process, Bibliometric Analysis, Science Mapping, Co-authorship analysis, Co-occurrence analysis

Conference Paper

Viktoria Paschalidou, Nikolaos Nousias, Kostas Vergidis (2022), “A Bibliometric Analysis on Business Processes: Preliminary Results”, *in 4th Summit on Gender Equality in Computing*, Thessaloniki, Greece, 16-17 June 2022

Acknowledgments

Upon the completion of my thesis I would like to thank the people that have supported me throughout this journey.

First and foremost I would like to wholeheartedly thank my Supervisor, Dr Konstantinos Vergidis, for trusting me from the beginning of our collaboration. His empathetic and supportive approach served as a lighthouse in times of doubt and was critical for reaching our end goal. Transmitting his love for research and including me in his team were also precious things that made me see the field of research from a different angle and rekindled my interest in contributing to the scientific community further. I feel very grateful for this collaboration.

At the same time, I would like to express my gratitude to my best friend and my family who have been supporting my research aspirations and the decisions they entail from the beginning until the end. Besides the things I learned in the field of my research, I learned a lot about myself too, and that is because of them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

Since 2019 when the COVID-19 pandemic started the business arena has changed dramatically. Business processes have developed rapidly and much of the research done could be considered obsolete just within three years. Today, businesses have to be more agile than ever while ensuring sustainability.

Business processes are the cornerstone of the welfare of a business and the study of the trends in the scientific field is crucial, in order to foresee the direction that businesses should take.

1.1. Business processes

A business process is a sequence of predefined steps, performed with the aim to satisfy a business objective. Many definitions are found in the literature [Figure 1] and for the premise of this thesis, the definition proposed by Vergidis is adopted: "A business process is perceived as a collective set of tasks that when properly connected and sequenced perform a business operation. The aim of a business process is to perform a business operation, i.e., any service-related operation that produces value to the organization".

Organizations are doomed to fail when they are unaware of the business processes that they need to support. Therefore, they operate inefficient processes that thwart them from fulfilling the demands of today's changing world, while they continuously fail to adopt the rapidly changing technology and follow the global competition. Altogether, organizations which endeavour to survive in the long-term, need to operate responsive and adaptable processes, denoting in that way the importance of process flexibility and the overall focus on the business processes concept (Nousias, 2021).

The importance of business processes has become imperative in the most developed economies since the mid-eighties and today the whole contemporary organizational structure emphasizes the role of business processes (Lizano-Mora et al., 2021), with the adoption of development strategies and methods like the Business Process Management (BPM) and the Business Process Reengineering (BPR).

The significance of the aforementioned is substantiated by the great amount of business process-related studies. However, the global changes in the recent years can be considered milestones for businesses thus it is worth researching which are the current trends in the field.

Author(s)	Business Process Definition
Hammer and Champy	"A business process is a collection of activities that takes one or more kinds of input and creates an output that is of value to the customer."
Davenport	"A process is simply a structured, measured set of activities designed to produce a specified output for a particular customer or market."
Jacobson	A business process is "the set of internal activities performed to serve a customer."
Stohr and Zhao	"A business process consists of a sequence of activities. It has distinct inputs and outputs and serves a meaningful purpose within an organization or between organizations."
Weske	"A business process consists of a set of activities that are performed in coordination in an organizational and technical environment."
Vergidis	"A business process is perceived as a collective set of tasks that when properly connected and sequenced perform a business operation. The aim of a business process is to perform a business operation, i.e., any service-related operation that produces value to the organization".
Chinosi and Trombetta	"A business process (BP) is a set of one or more linked procedures or activities executed following a predefined order which collectively realize a business objective or policy goal, normally within the context of an organizational structure defining functional roles or relationships".

Figure 1. Business process definitions (Nousias, 2021)

1.2. Aim and objectives

The main aim of this thesis is to examine the publication output regarding “business processes” until 2022. The subsequent objectives are

- To find an appropriate review method which is able to produce credible and objective results while taking into consideration all the existing literature
- To examine the degree to which the field is shaped by diverse contributions and multidimensional aspects
- To discover the trends in the field which can reflect the major areas of future development

1.3. Thesis layout

The rest of the thesis is structured as follows. Chapter 2 begins with a comparison of the main review methods and Bibliometric Analysis is the selected one, in alignment with the aim of the study. Bibliometric Analysis is further analyzed and the steps of its implementation are considered. A complementary preprocessing procedure is also proposed. Next, in Chapter 3 the proposed Bibliometric Analysis preprocessing is deployed for the aim of this study and a first examination of the available bibliometric techniques, the field of research, the bibliographic databases and the software is performed. Then in Chapter 4, the Bibliometric Analysis process acquired from the bibliography is deployed and the final selection of the aim, the bibliometric techniques, the search query, the bibliographic databases and the software is performed. Chapter 5 exhibits the results of the Bibliometric Analysis and Chapter 6 provides the thesis overview and the research contribution, the research limitations, as well as establishes directions for future work, stemming from this research.

2. Bibliometric Analysis

This chapter starts with a presentation of the characteristics of Bibliometric Analysis in comparison with two prominent review methods: Meta-analysis and Systematic Literature Review. Afterwards, the main techniques of the Bibliometric Analysis are introduced and ultimately the procedure of the analysis is put under the microscope.

Bibliometric Analysis is one of the three major review methods, besides Meta-analysis and Systematic Literature Review. Their main differences appear in [Figure 2]. In essence, Meta-analysis is often used as a theory extension tool. It is quantitative in nature and is able to handle large amounts of literature, providing a nuanced summary of a given field. However, the literature considered tends to be less diverse, and the heterogeneity of existing studies and the existence of a publication bias can have an adverse effect on the validity of the results obtained. On the other hand, Systematic Literature Review is better suited for confined or niche research areas. It is qualitative in nature, and requires a narrow scope of study and thus includes a little number of papers for review.

Bibliometric Analysis encapsulates the application of quantitative techniques (e.g. citation analysis) on data, which tend to be massive and objective in nature (e.g. number of citations, occurrences of keywords). By making sense of the large volume of data, Bibliometric Analysis aims to decipher and map the cumulative scientific knowledge of a specific domain in the existing literature. The exploration of the intellectual structure of a research field can exhibit publication and collaboration patterns, reveal research constituents - which can lead to the identification of knowledge gaps, and uncover emerging trends - which can lead to the discovery of novel ideas for investigation.

The proliferation of bibliometrics is relatively new in the business research field, with a growing tendency of relevant studies over the past years, in alignment with the overall growth of the scientific research itself. The need for this kind of analysis emerges from the fact that large bibliographic datasets have made the other review methods cumbersome and impractical (Donthu et al., 2021). The recency of the deployment of Bibliometric Analysis in business research results in a deficiency of reliable and extensive guides. Consequently, it is challenging for scholars to learn more about the methodology itself and its application in the field. Papers like (Donthu et al., 2021) provide an overview of the bibliometric methodology and some steps for conducting the analysis, however, the step-by-step process of the analysis is not addressed in

detail. This is understandable, because, on the one hand, various software can be used for the data analysis (Fahimnia et al., 2015) (Donthu et al., 2021) (Sajovic & Boh Podgornik, 2022) (Lizano-Mora et al., 2021) (Manzari, 2021), such as

- BibExcel
- CiteSpace
- Gephi
- HistCite
- Leximancer
- Pajek
- PoP (Publish or Perish)
- RStudio and R Bibliometrix package
- SciMAR
- Sci2
- SITKIS
- UCINET
- VOSviewer

and, on the other hand, different bibliometric techniques exist; the main ones being Performance Analysis and Science Mapping [Figure 3].

Essentially, Performance Analysis accounts for the contributions of research constituents, whereas Science Mapping focuses on the relationships between the research constituents. Among them, Performance Analysis is the most prominent one, and reviews that deploy it showcase the performance of different research constituents (e.g. authors, institutions, countries, and journals) in the field under study. However, in the context of Bibliometric Analysis, many studies do not make full use of its potential, either by relying on limited datasets or limited techniques (e.g., Performance Analysis without Science Mapping). Such limitations inevitably result in partial understanding of a field (Donthu et al., 2021).

From the aforementioned emerges the need for an extensive guide that will (1) expound on Science Mapping techniques in a comprehensive manner and (2) provide a step-by-step process for the actual data analysis while orienting the application in the business field.

2.1. Bibliometric Analysis approach

As discussed earlier, it can be challenging for a new researcher to find a step-by-step guide for the deployment of Bibliometric Analysis in the business field. Different sources provide different chunks of information and one should put effort into combining all the pieces together. What, unquestionably, supports understanding is visualization and the majority of papers do not include flowcharts of the procedures that were followed. Some of the business-oriented papers that include visualization of the Bibliometric Analysis procedure are: (Ali et al., 2022) exhibit a search flow diagram, which is following the PRISMA guidelines, in order to screen the documents for analysis. (Donthu et al., 2021) provide a flowchart of the Bibliometric Analysis procedure and a toolbox of the available techniques. (Sajovic & Boh Podgornik, 2022) and (Lizano-Mora et al., 2021) utilize a workflow dedicated to Science Mapping.

The present work was based on the flowchart of (Donthu et al., 2021), which was considered holistic but at the same time thorough and comprehensive. The suggested Bibliometric Analysis procedure has four steps: it starts with the definition of the aim and scope of the study, continues with the selection of the technique, then data collection follows, and concludes with the actual analysis. The flowchart comprises clear steps and the paper proposes additional questions, that support the process of its application. [Figure 4] is the result of merging the given questions with each step of the procedure.

In the process of applying the steps for the purpose of this study, it became clear that reiterations of the first three steps were necessary. Starting from the scope, it became clear that one cannot foresee the scope (Step 1), without having an overview of the research work available (Step 3). Another realization was that the aim of a study (Step 1) is highly influenced by the limitations of the Bibliometric Analysis itself (Step 2). Since each bibliometric technique produces certain outputs this factor should be taken into account beforehand, in order to define the aim of a study. Then, the choice of a technique (Step 2) is interconnected with the availability of data (Step 3). Ultimately, different databases and different software for data analysis pose additional limitations, which might result in more reiterations, with consequent modifications of the aim or reconsiderations of the techniques. For this reason, an additional preprocessing procedure is proposed [Figure 5], before the initiation of the main procedure. The additional approach aims to decrease the reiterations needed, through realizations that occur upon experimentation.

Review type	Goal	When to use	When not to use	Scope	Dataset	Analysis
Bibliometric analysis	• Summarizes large quantities of bibliometric data to present the state of the intellectual structure and emerging trends of a research topic or field.	• When the scope of review is broad. • When the dataset is too large for manual review.	• When the scope of review is specific. • When the dataset is small and manageable enough that its content can be manually reviewed.	• Broad	• Large	• Quantitative (evaluation and interpretation) • Qualitative (interpretation only)
Meta-analysis	• Summarizes the empirical evidence of relationship between variables while uncovering relationships not studied in existing studies.	• When the focus of review is to summarize results rather than to engage with content, which may be broad or specific. • When studies in the field are homogenous. • When the number of homogeneous studies available is sufficiently high. • When the number of homogeneous studies remaining after removing low quality studies is sufficiently high.	• When studies in the field are heterogeneous. • When the number of homogenous studies is relatively low. • When the number of high-quality homogeneous studies is relatively low.	• Broad • Specific	• Large • Small but adequate	• Quantitative (evaluation and interpretation)
Systematic literature review	• Summarizes and synthesizes the findings of existing literature on a research topic or field.	• When the scope of review is specific. • When the dataset is small and manageable enough that its content can be manually reviewed.	• When the scope of review is broad. • When the dataset is too large for manual review.	• Specific	• Small	• Qualitative (evaluation and interpretation)

Figure 2. Comparison of three major review methods (Donthu et al., 2021)

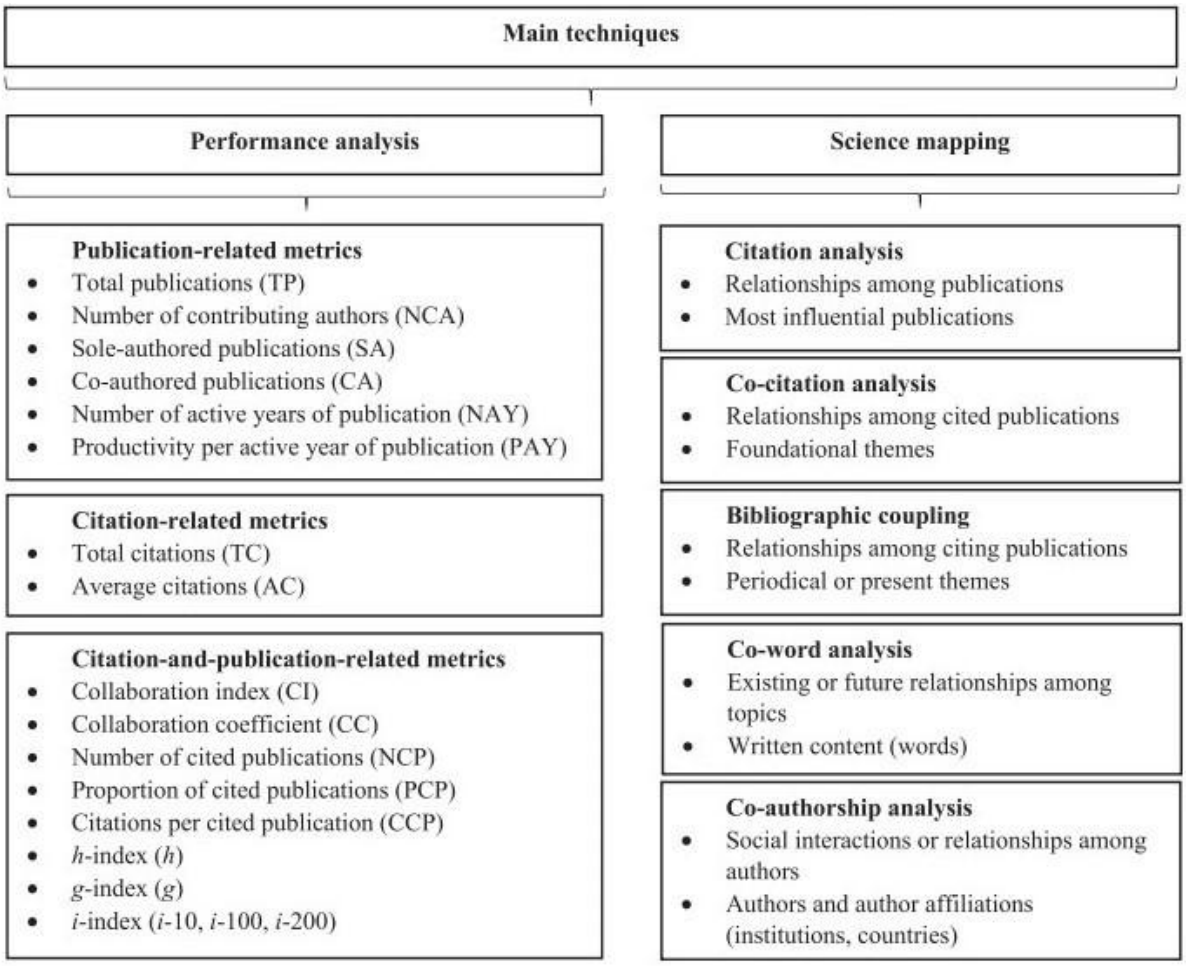


Figure 3. Bibliometric Analysis main techniques (Donthu et al., 2021)

The proposed procedure starts with a profound understanding of the Bibliometric Analysis techniques (Performance Analysis and Science Mapping techniques). Understanding the main principles and assumptions behind each technique (e.g. Co-authorship analysis) sheds light on which data is required in each case and which potential outputs can be produced.

While keeping this in mind, one can proceed with the exploration of the available sources, which fall under the research interest. At this point, it is still under question whether the Bibliometric Analysis is necessary. After experimentations with queries, it may become clear that, despite the fact that the scope becomes broad, there is lack of a high amount of available data (i.e. not many publications). In this case, a first consideration for another review method should be given.

Afterwards, the limitations of various databases should be investigated. It is highly likely that different databases include different research work, thus the former queries should be utilized again, to confirm or disprove the consideration made in the previous step. At this point, the researcher is in a good position to decide whether Bibliometric Analysis should be deployed.

Simultaneously, the experimentations with queries and the observations on the availability of data, gradually indicate the most suitable scope of the study. The last, critical thing in this step, is the exploration of the limitations with regard to the export options that each database offers, both in terms of the bibliographic data that can be exported (e.g. author keywords) and the format of the output file (e.g. CSV, BibTeX). The limitations are critical for the final decision on the bibliometric techniques that can be deployed.

Lastly, experimentation with different software packages is essential. Different software supports different imported files, thus, in combination with the previous step, this should be a primary thing to be taken into account. Running a small sample of data can give a first impression of both the complexity of the software and its potential (visual) output. Potential limitations with the exported file format should also be examined.

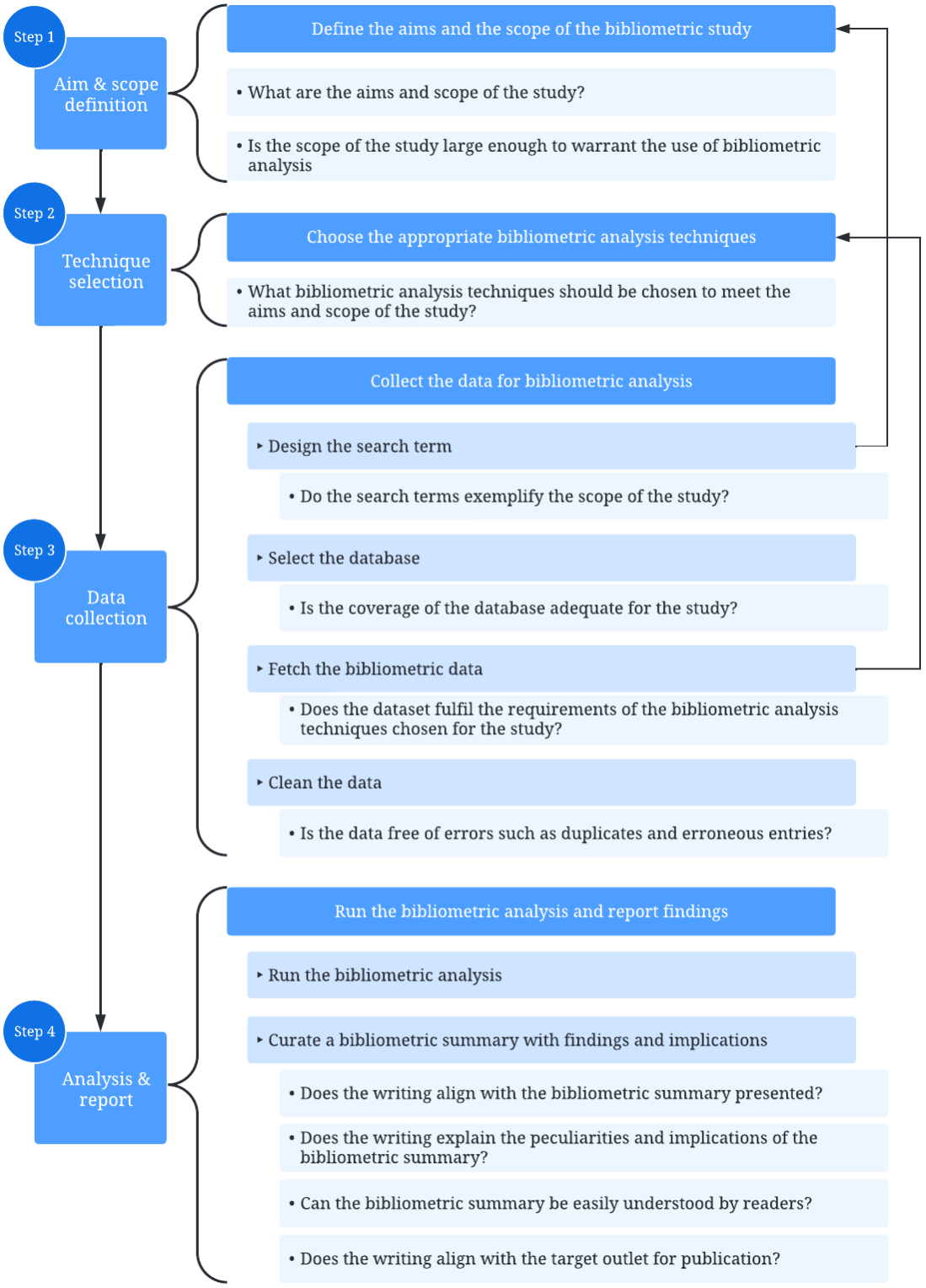


Figure 4. Bibliometric Analysis procedure (Donthu et al., 2021)

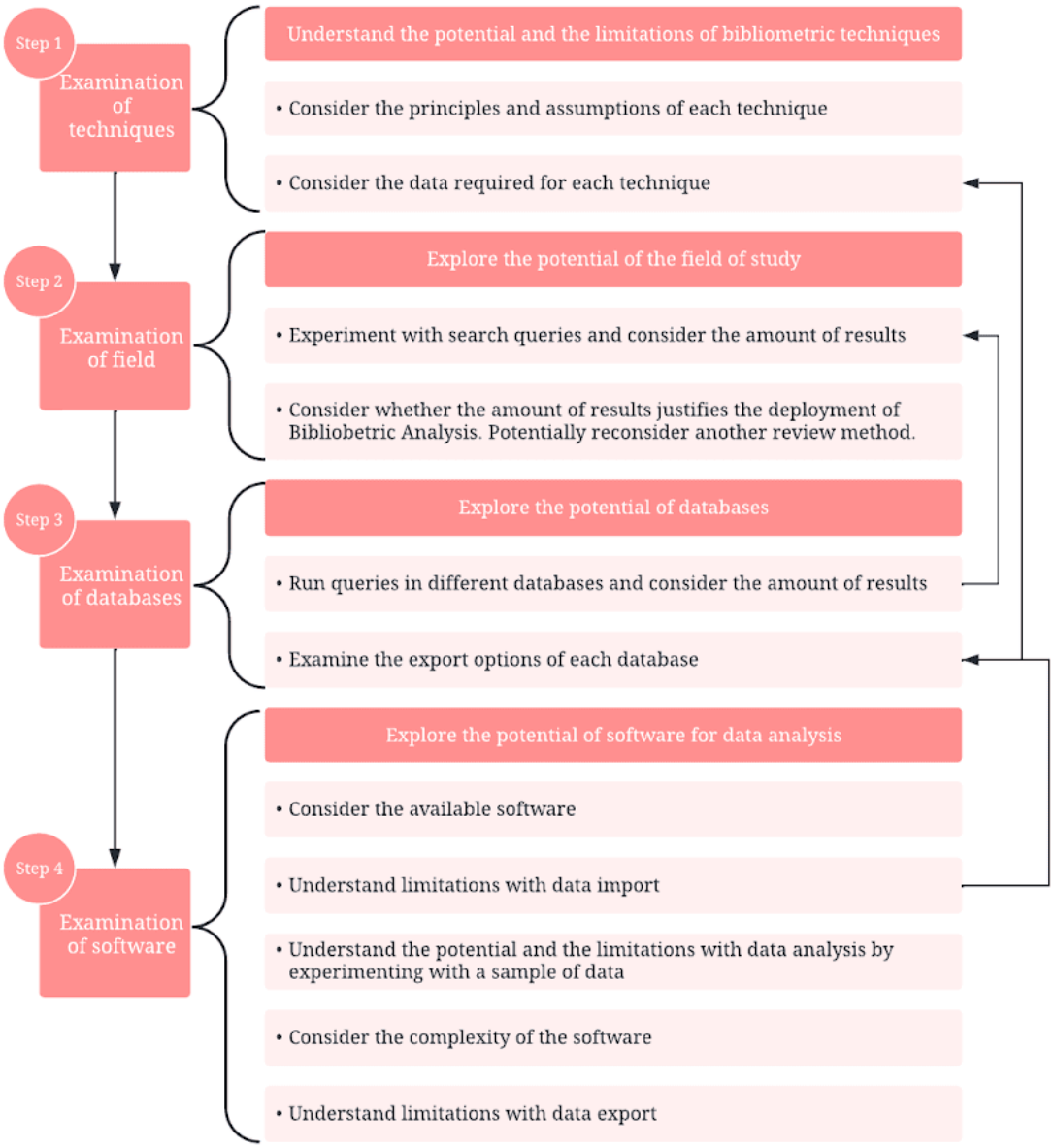


Figure 5. Proposed Bibliometric Analysis procedure

After going through this empirical process, a researcher can confidently initiate the core Bibliometric Analysis procedure. The scope and aim can be defined with a higher degree of certainty and the techniques can be chosen while keeping in mind all the limitations that databases and software pose. Hence, the subsequent steps of data collection and analysis can be the main focus of the researcher without facing critical, unforeseen obstructions that will cause more reiterations. The proposed, experiential preprocessing procedure, has been the backbone of this study and is analytically presented in the next chapter.

3. Stage 1: Bibliometric Analysis preprocess

This chapter aims to provide a step-by-step guide on the application of the proposed Bibliometric Analysis preprocess. Each step of the process is analyzed and deployed while showcasing the progression of this study.

3.1. Step 1: Examination of techniques

The goal of the first step of the Bibliometric Analysis preprocess is to understand the potential and the limitations of the bibliometric techniques. Performance Analysis has been widely used since its metrics can be easily understood and computed for each research constituent (e.g., authors, institutions, countries, journals) either as an aggregate (per research constituent) or in specific (e.g., research constituent per publication, per year, or per period) (Donthu et al., 2021). Thus in the current study effort was put into deconstructing the Science Mapping techniques.

[Table 1], [Table 2], [Table 3], [Table 4], and [Table 5] provide an overview of the assumptions, interpretation, and outcomes of the main Science Mapping techniques: Co-authorship, Co-occurrence, Citation, Bibliographic coupling, and Co-citation analysis, respectively (Donthu et al., 2021) (Donthu et al., 2020) (Sajovic & Boh Podgornik, 2022) (Ortigueira-Sánchez & Risco-Martínez, 2021).

From the tables, it becomes clear that the data required for the analysis are mainly: authors' names and their affiliations (i.e. countries, institutions), words (e.g. keywords), references and citations.

Co-authorship Analysis

Assumption			
The relationship of items (authors, countries, institutions) depends on the number of documents co-authored.			
Unit of Analysis	Analysis	Interpretation	Outcomes
Authors	Social interactions or relationships among scholars	Researchers are linked to each other based on the number of publications they have authored jointly.	• Intellectual associations among scholars • Collaboration patterns • Collaboration mapping across different time periods • Impact of authors' relationships on the development of the research field
Author Affiliation: **Countries**	Intellectual associations among scholars in different countries	When authors from two or more countries contribute to a given article, the authors' countries are considered as collaborating countries. Countries are linked to each other based on the number of publications they have authored jointly.	• Volume of joint papers in each country • Collaboration intensity of countries • Clusters of scholars from particular regions • Impact of countries' relationships on the development of the research field
Author Affiliation: **Institutions**	Intellectual associations among scholars in different institutions	Research institutions are linked to each other based on the number of publications their authors have authored jointly.	• Impact of institutions' relationships on the development of the research field

Table 1. Co-authorship analysis overview

Co-occurrence analysis

Assumption			
Words that frequently appear together have a thematic relationship with one another.			
Unit of Analysis	Analysis	Interpretation	Outcomes
Words (Title, Abstract, Author keywords, Index keywords, Full text)	Written content of publication	Association among keywords The relationship of items depends on the number of documents in which they appear together	• Most popular keywords used by researchers • Conceptual/knowledge structure of the literature • Recognition of main research areas in the field • Revelation of existing or future relationships among topics in the research field
Limitations			
Some words are used in multiple contexts thus (re)reading of publications might be necessary. *Some words can be very general and it can be challenging to be assigned to clusters.*			

Table 2. Co-occurrence analysis overview

Citation analysis

Assumption			
Citations reflect intellectual linkages between publications that are formed when one publication cites another. Thus, the impact of a publication is determined by the number of citations it receives.			
Unit of Analysis	Analysis	Interpretation	Outcomes
Documents (Author name, Citations, Title, Journals, DOI, References)	Relationships among publications	The relationship of items depends on the number of times they cite each other	• Understanding of intellectual dynamics of the field • Most influential publications in the research field

Table 3. Citation analysis overview

Bibliographic coupling

Assumption			
Scientific works exhibit intellectual convergence on the basis of their common sources and patterns of referencing. When two articles cite a common third article, this implies that both articles deliberate on and stress similar discussions. In other words, two publications sharing common references are also similar in their content.			
Unit of Analysis	Analysis	Interpretation	Outcomes
Documents (Author name, Title, Journals, DOI, References)	Relationships among citing publications	Clusters of publications are thematic The relationship of items depends on the number of references they share	• Understanding of periodical or present development of themes in research field

Table 4. Bibliographic coupling Analysis overview

Co-citation analysis

Assumption			
Frequent citation of two or more references in a third document exhibit conceptual or intellectual similarities among the citing and cited documents. That means that when two publications co-occur (i.e. are co-cited) in the reference list of another publication they are connected thematically.			
Unit of Analysis	Analysis	Interpretation	Outcomes
Documents (References)	Relationships among cited publications	The relationship of items depends on the number of times they have been cited together The larger the number of publications by which two publications are co-cited, the stronger the co-citation relation between the two publications Clusters of publications are thematic	• Intellectual structure of research field • Discovery of most influential publications • Semantic similarities of publications • Understanding of development of foundational themes in research field
Limitations			
Concentration on highly-cited publications leaves recent or niche publications out of the thematic clusters.			

Table 5. Co-citation analysis overview

3.2. Step 2: Examination of the field

Upon comprehension of the Science Mapping techniques and the required data for their deployment, the next step is to experiment with search queries in order to get a first estimation of the number of available publications. The "Scopus" bibliographic database was chosen arbitrarily, with the only criterion of familiarity with its functionalities. The data was accessed on 04.06.2022 and the terms "BPM" and "Business process management" were used as the main keywords of the search, since they were found to be representative of the field. The outcomes of the experimentations are illustrated in [Table 6]. Besides alternations of the operators "OR" and "AND", trials were performed with regard to the document types and publication years. Additionally, the search was limited to publications authored in English.

SCOPUS

Query	Results
TITLE ("bpm") AND (LIMIT-TO (LANGUAGE , "English"))	1.594
TITLE ("business process management") AND (LIMIT-TO (LANGUAGE , "English"))	1.371
TITLE ("bpm") OR TITLE ("business process management") AND (LIMIT-TO (LANGUAGE , "English"))	2.813
TITLE ("bpm") OR TITLE ("business process management") AND (LIMIT-TO (SRCTYPE , "p") OR LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (LANGUAGE , "English"))[1]	2.141
TITLE ("bpm") OR TITLE ("business process management") AND (LIMIT-TO (SRCTYPE , "p") OR LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (EXACTKEYWORD , "Business Process Management") OR LIMIT-TO (EXACTKEYWORD , "BPM"))	909
TITLE ("bpm") OR TITLE ("business process management") AND PUBYEAR > 1999 AND (LIMIT-TO (SRCTYPE , "p") OR LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (EXACTKEYWORD , "Business Process Management") OR LIMIT-TO (EXACTKEYWORD , "BPM"))	902
TITLE ("business process management") AND PUBYEAR > 2000 AND (LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "ch") OR LIMIT-TO (DOCTYPE , "re")) AND (LIMIT-TO (LANGUAGE , "English"))[2]	1.173

Table 6. Scopus search results

[1] Source Types: Conference Proceeding (p) / Journal (j)

[2] Document types: Conference Paper (cp) / Article (ar) / Book Chapter (ch) / Review (re)

The outcome of the search indicated that Bibliometric Analysis can be justified, with the highest amount of results being 2.813 when the query *TITLE ("bpm") OR TITLE ("business process management") AND (LIMIT-TO (LANGUAGE , "English"))* was used.

3.3. Step 3: Examination of databases

To explore the possibility of acquiring even more documents for analysis, further examination of bibliographic databases was conducted. Google Scholar was chosen, due to its open accessibility, and Emerald Insight, Science Direct, and IEEE Xplore were chosen due to institutional accessibility. The queries used were similar to the ones used in Scopus, with differences in the syntax that each database required. The [Table 7], [Table 8], [Table 9] and [Table 10] exhibit the queries which ran for each database, together with the number of results that were retrieved. The data was accessed on 04.06.2022.

GOOGLE SCHOLAR

Query	Results
allintitle: "BPM"	12.700
allintitle: "business process management"	5.870
allintitle:"BPM" OR allintitle:"business process management"	13.400
allintitle: "BPM" OR "business process management"	17.600
allintitle: "BPM" OR "business process management" (with filter 2000-2022)	15.400

Table 7. Google Scholar search results

EMERALD INSIGHT

Query	Results
title:"BPM"	53
title:"business process management"	157
title:"BPM" OR (title:"business process management")	139

Table 8. Emerald Insight search results

ScienceDirect

Query	Results
Title: "business process management"	96
Title: "bpm"	102
Title: "bpm" OR "business process management"	193

Table 9. ScienceDirect search results

IEEE Xplore

Query	Results
"Document Title":"business process management"	224
"Document Title":"bpm"	281
("Document Title":"bpm") AND ("Author Keywords":"bpm")	77
("Document Title":"bpm") OR ("Document Title":"business process management")	498
("Document Title":"bpm") OR ("Document Title":"business process management") AND ("Author Keywords":"bpm")	322
("Document Title":"bpm") OR ("Document Title":"business process management") AND ("Author Keywords":"bpm") with filter "Publication topics" • business data processing • business process re-engineering • organisational aspects • educational institutions	131

Table 10. IEEE Xplore search results

The results indicated that Google Scholar provides the highest amount of results, with the highest number being 17.600 documents when *allintitle: "BPM" OR "business process management*" was used as query. However, it was observed that Google Scholar did not seem to have any export option [Figure 6]. With some further investigation, it was confirmed that Google Scholar does not offer a bulk download option. The only way to download bibliographic data is for the user to "Star" the documents manually (one by one) so that they are added to the "Library". Then, for the documents added to the "Library" the export option for bibliographic data

is available. With this in mind, Google Scholar was excluded from the potential databases that would be used for the study due to its incapacity to easily export the dataset.

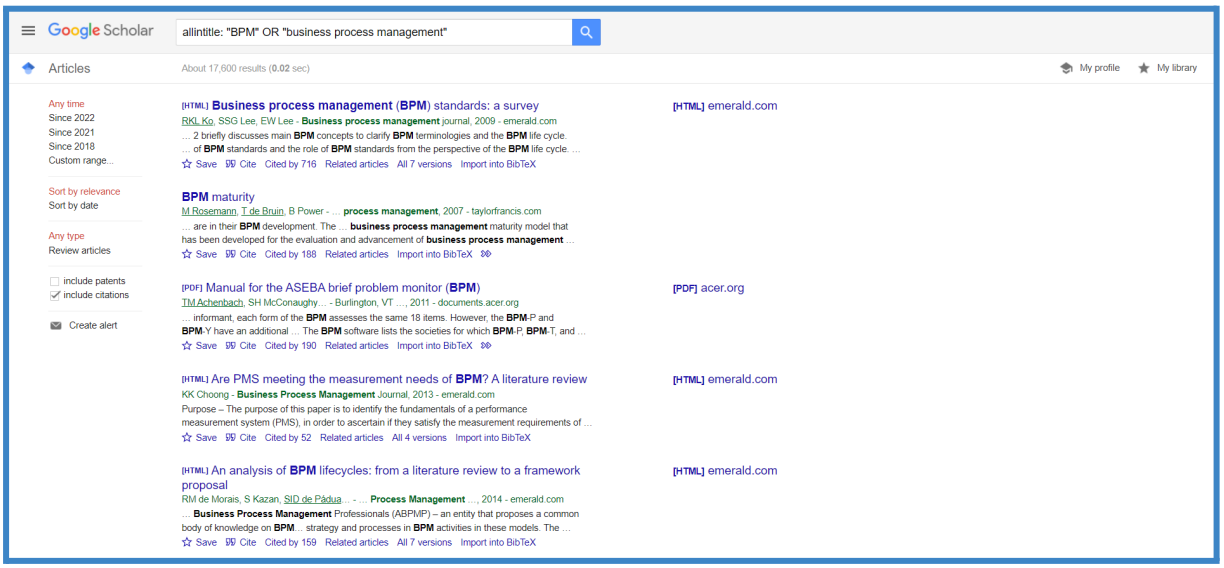


Figure 6. Google Scholar interface

For the rest of the databases, the option to download the results in batches was available and various export formats were supported: ScienceDirect and IEEE Xplore offer “Plain Text”, “BibTeX”, “RIS” and “RefWorks” export options while Emerald Insight only offers the “RIS” export option. Nevertheless, these datasets were also considered ineligible, since the number of retrieved publications was insufficient for the deployment of Bibliometric Analysis.

Another interesting observation was that when *"Document Title":"bpm"* was used as a query in the IEEE Xplore database, the top “Publication Topics” which appeared in the filters were

- business data processing (71)
- particle beam diagnostics (61)
- storage rings (34)
- electron accelerators (32)
- optical waveguide theory (28)

and with a quick search it was discovered that “BPM” also stands for

- beam position measurement (BPM)
- beam propagation method (BPM)
- beam position monitor (BPM)

thus, it was decided to exclude the term from the final query choice in order to retrieve more relevant results.

Ultimately, it was noticed that, in all the databases and queries, the significant majority of "Publication/Document types" are "Journals", "Articles" and "Conference Papers" and that the "Books", "Magazines", and other document types' ratio does not affect the results. Therefore, it was decided to dismiss the limitation concerning the "publication type" in the final search query.

This step ended with the deliberation of Scopus as the database, which seemed the most promising for providing the final dataset, which would consist of a big amount of data, required for the Bibliometric Analysis. Also, it was decided to neglect the "BPM" term and the limitations regarding "publication type" from the final query.

3.4. Step 4: Examination of software

In the next step, the potential of the available software was examined.

3.4.1. Data analysis software

As mentioned by (Donthu et al., 2021) Science Mapping techniques, when combined with network analysis, are instrumental in presenting the bibliometric structure and the intellectual structure of the research field. More specifically, (Fahimnia et al., 2015) support that "network analysis through bibliometric tools can prove powerful for identifying established and emerging topical areas. It can also help identify the clusters of research and researchers showing how the various areas of thought may have emerged based on author and institutional characteristics. Identifying the more influential researchers within the clusters sets the stage for determining additional emergent study fields through capturing more recent topics covered by these researchers".

Recommendations for software were found in publications (see Chapter 2). Initially, BibExcel (Persson, n.d.) was selected due to its prominence in the publications. From the very beginning, the obsolete interface of BibExcel [Figure 7] impeded the immediate understanding of its functionalities and after much experimentation, it was concluded that its overall operating environment was more complex than expected. In fact, the manual of the software (Persson et al., 2009) elaborated mainly on the import and export functionalities and the supplementary audiovisual documentation was not in English but in Spanish. Besides the official documentation, the guides and video tutorials that were additionally found on the internet were limited and did not extensively present all the features of the software. What was additionally realized was that BibExcel does not provide visual outputs but rather produces network files (NET-file) and vector

files (VEC-file) that can be imported into other software like Pajek. [Figure 8], [Figure 9], [Figure 10], and [Figure 11] showcase the process for the extraction of the publication year frequencies for a sample of publications.

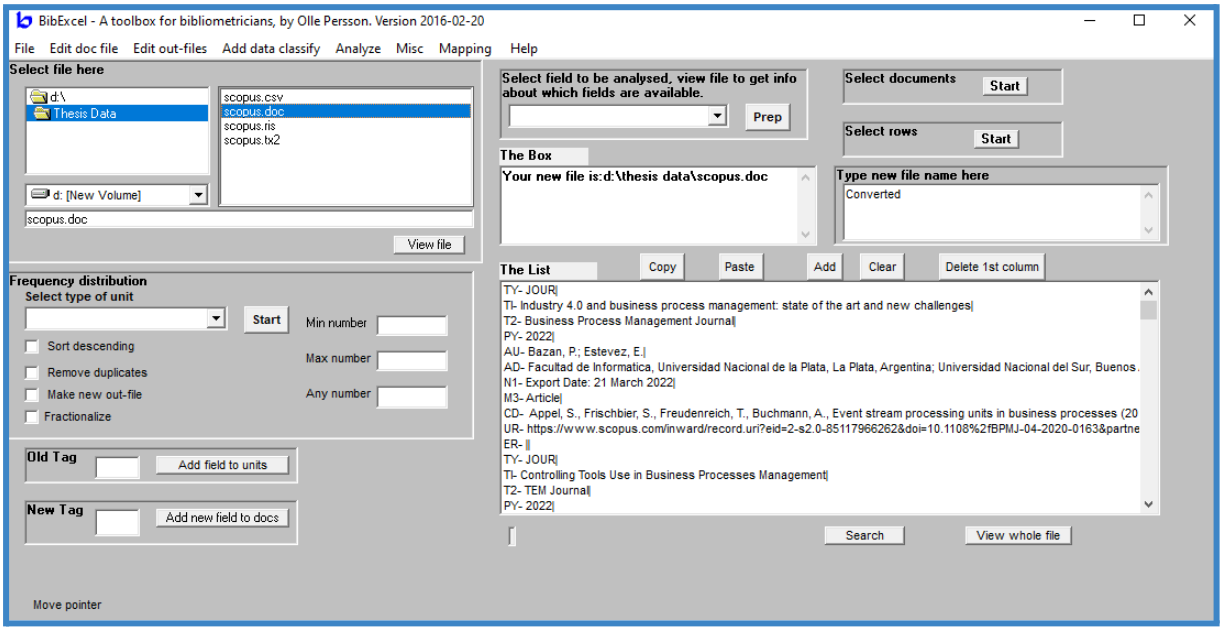


Figure 7. BibExcel interface

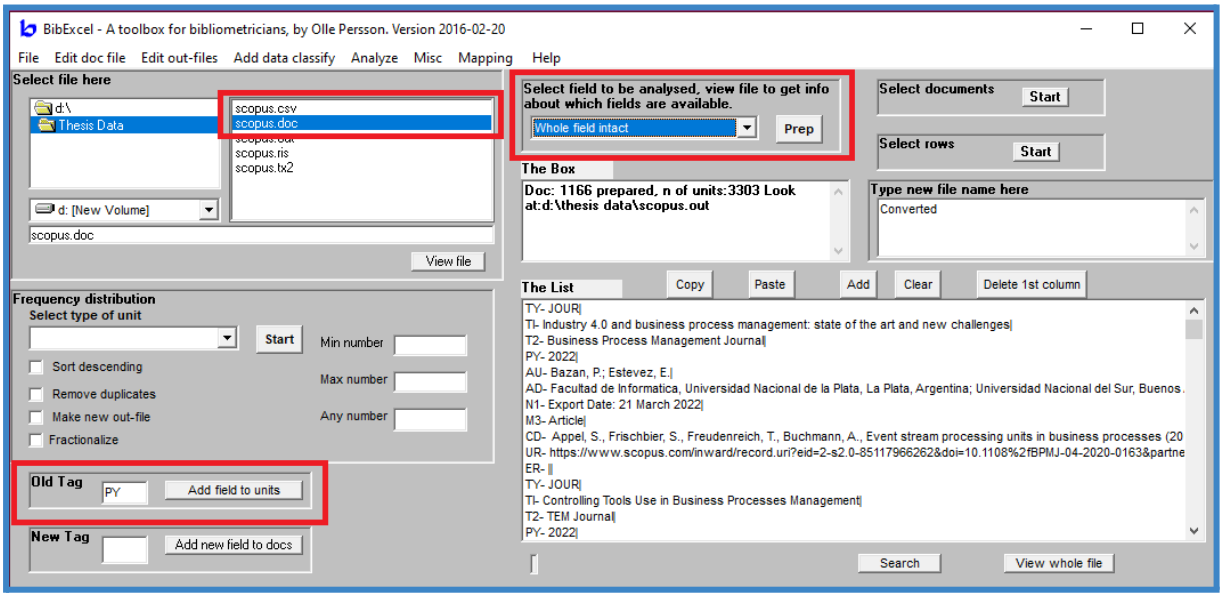


Figure 8. BibExcel: Parameter setting for finding the publication year

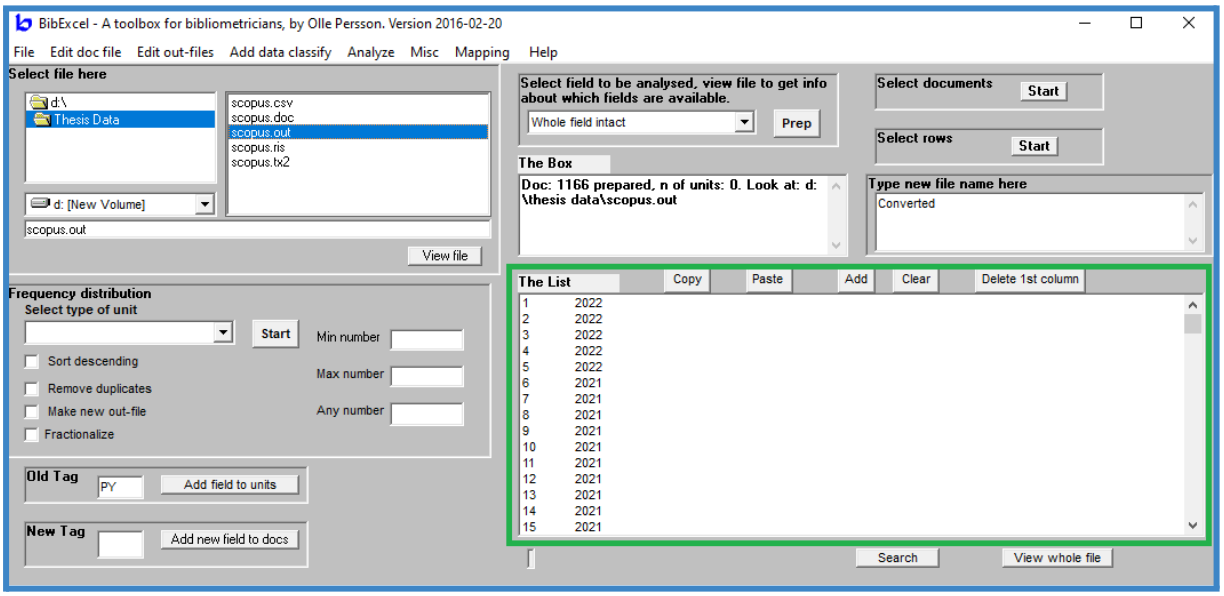


Figure 9. BibExcel: Output of publication year of each document

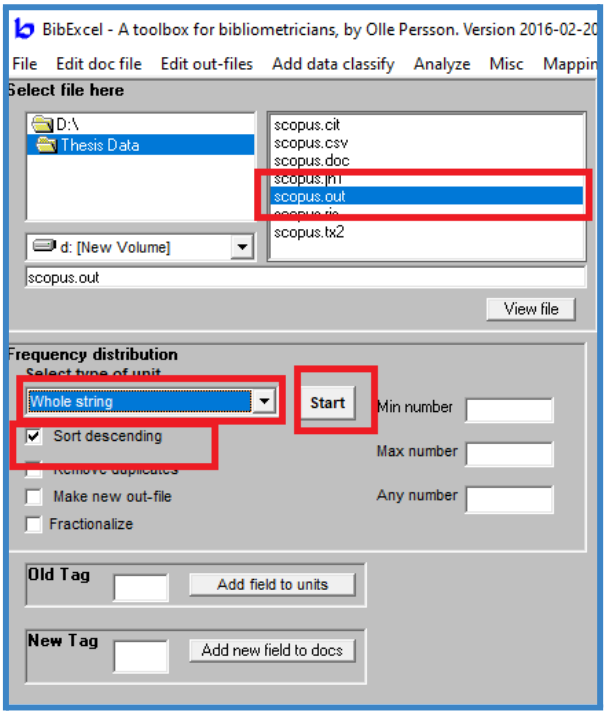


Figure 10. BibExcel: Parameter setting for finding frequencies of each publication year

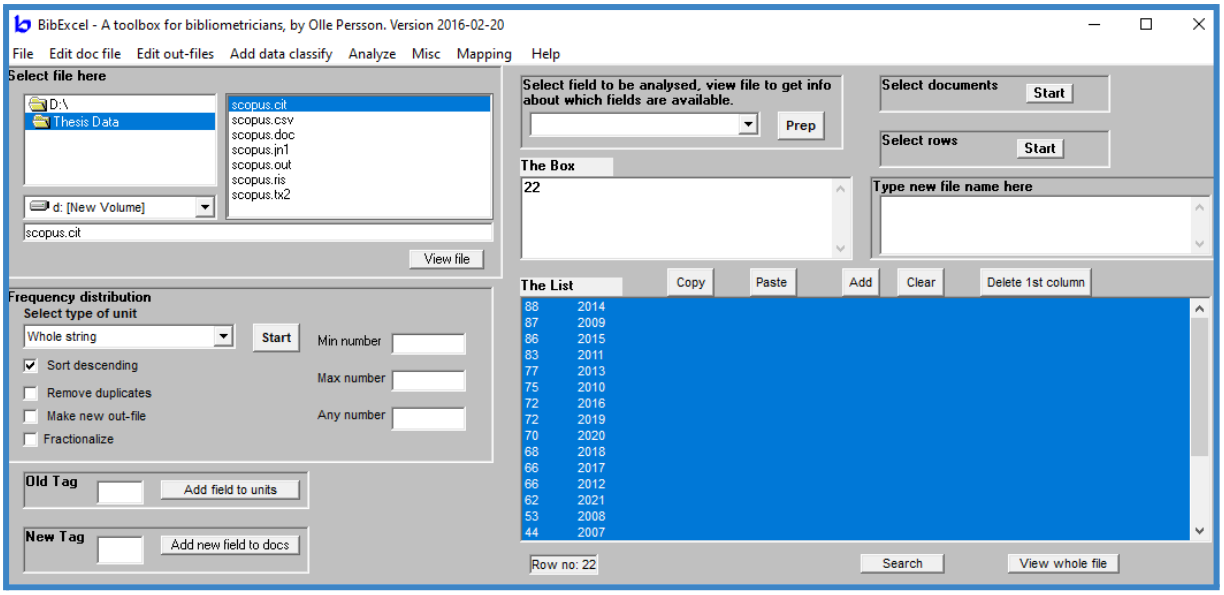


Figure 11. BibExcel: Output of frequency of publication years

For all the aforementioned reasons, it was decided to abandon BibExcel and search for another software, based on the criterion of the availability of thorough guides and video tutorials that would support the quick understanding of its main functionalities. VOSviewer (VOSviewer, 2022) was assessed to be more simple, yet promising software to experiment with. VOSviewer is a software tool for constructing and visualizing bibliometric networks. These networks may, for instance, include journals, researchers, or individual publications, and they can be constructed based on citation, bibliographic coupling, co-citation, or co-authorship relations. VOSviewer also offers text mining functionality that can be used to construct and visualize co-occurrence networks of important terms extracted from a body of scientific literature (Van Eck & Waltman, 2022).

VOSviewer supports the import of three types of reference manager files: RIS, EndNote, and RefWorks files as well as five types of bibliographic database files: Web of Science files, Scopus files, Dimensions files, Lens files, and PubMed files [Table 11]. In terms of the exported files, they can be in the format of a VOSviewer map, a VOSviewer network, JSON, GML, and Pajek. Additionally, screenshots can be saved in several graphic file formats, like PNG, EPS, PDF, and SVG, with the advantage that they can be resized without loss of quality. On top of all, with some experimentations, the interface of VOSviewer was regarded as user-friendly, and its visual output was highly interactive and suitable for the understanding of the relationships that Scientific Mapping methods indicate [Figure 12].

Thus, VOSviewer was selected as the software for conducting the Bibliometric Analysis.

Database	Website
Web of science	www.webofscience.com
Scopus	www.scopus.com
Dimensions	www.dimensions.ai
Lens	www.lens.org
PubMed	pubmed.ncbi.nlm.nih.gov

Table 11. VOSviewer bibliographic database files support

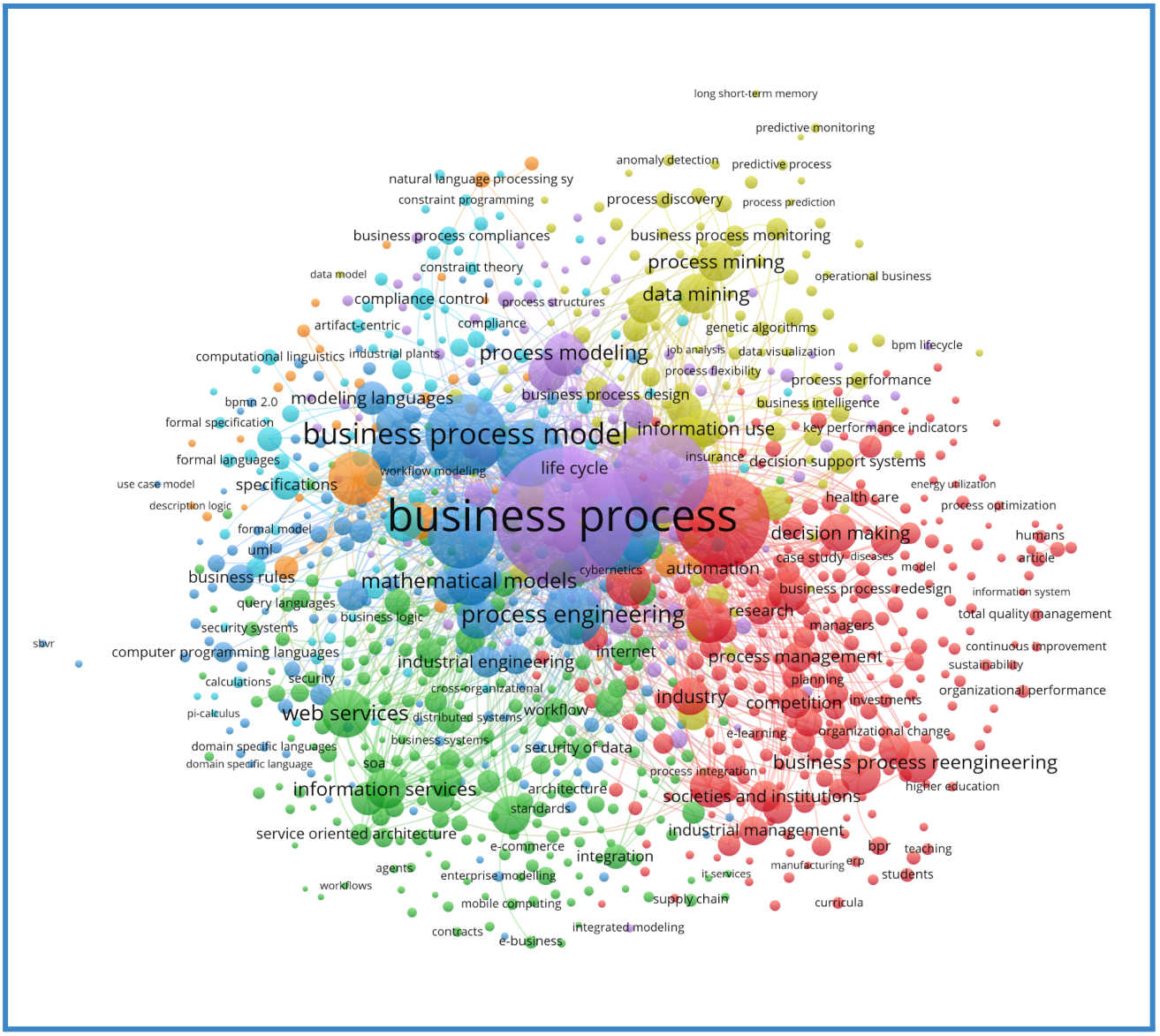


Figure 12. VOSviewer visual output

3.4.2. Data cleansing software

Besides the software for data analysis, software for data cleansing was additionally considered. Publications acquired from prominent databases, like Scopus, tend to have a high degree of coherency and it was assumed that most of the data required for the analysis would have had a unified formulation (authors' names, countries, references). However, the keywords of each publication are unique, and up to the author to define. Thus, differences with regard to their morphology were expected to be found. However, no references to software cleansing were found in the bibliography.

An open-source data cleansing software was searched on the internet and the first result that occurred suggested the OpenRefine software [Figure 13]. OpenRefine (previously known as Google Refine) is an open-source powerful tool for working with messy data: cleaning it, transforming it from one format into another, and extending it with web services and external data (OpenRefine, 2012). Among others, OpenRefine incorporates two features that were considered helpful for the data cleaning process: data exploration features and data transformation features. With the former feature, one can learn more about the dataset with sorting, filtering, and viewing options. Compared to spreadsheets, OpenRefine does not store formulas and display the output of those calculations; it only shows the value inside each cell. The latter feature addresses the cleaning and correction of the data. Unlike spreadsheets which need typing of formulas inside cells, OpenRefine automatically fixes typos and converts things to the right format. The main aspects of this feature are:

- change the order of rows or columns
- edit cell contents within a particular column
- transform rows into columns, and columns into rows
- split or join columns
- convert rows of data into multi-row records

The documentation of OpenRefine (OpenRefine User Manual, 2022) was very comprehensive and experimentations with a sample dataset were easy and effective.

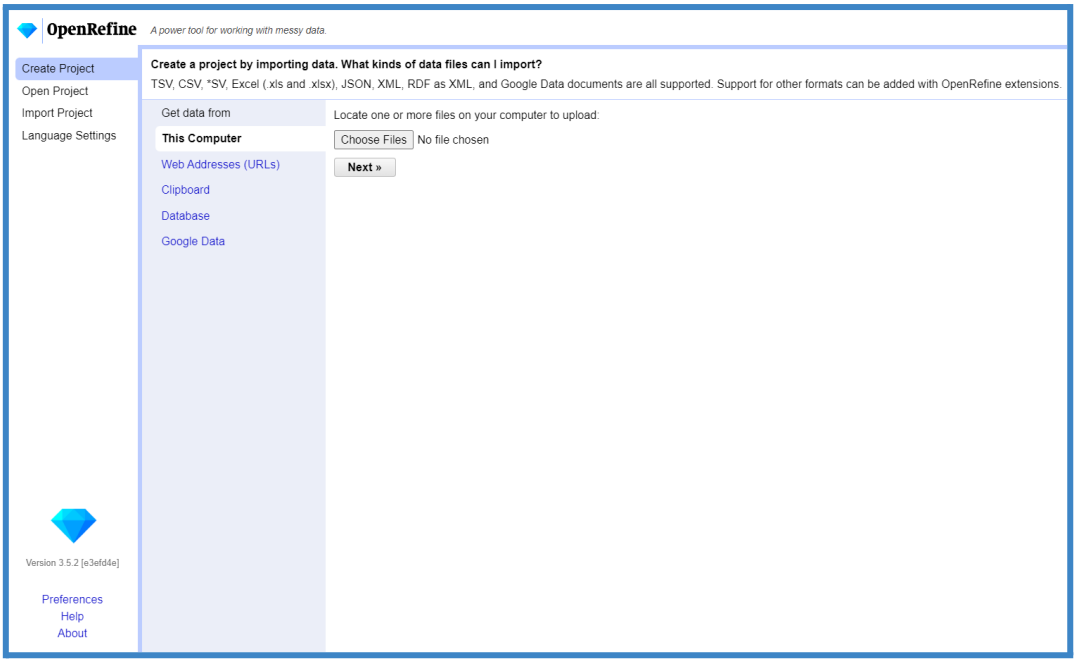


Figure 13. OpenRefine interface

Ultimately, the utility of OpenRefine was revisited with regard to the functionalities of VOSviewer. In terms of data exploration, VOSviewer primarily provides a visualization of the bibliometric data, and information about frequencies cannot be easily extracted (e.g. frequency of publication years). Thus, the data exploration feature of OpenRefine was considered valuable [Figure 14]. Additionally, VOSviewer integrates the identification of duplicate entries, however, it does not incorporate morphological inspection. Thus, the functionality of OpenRefine about data transformation was also considered beneficial. For these reasons, OpenRefine was considered supplementary to VOSviewer and it was decided to be the software for the data cleansing.

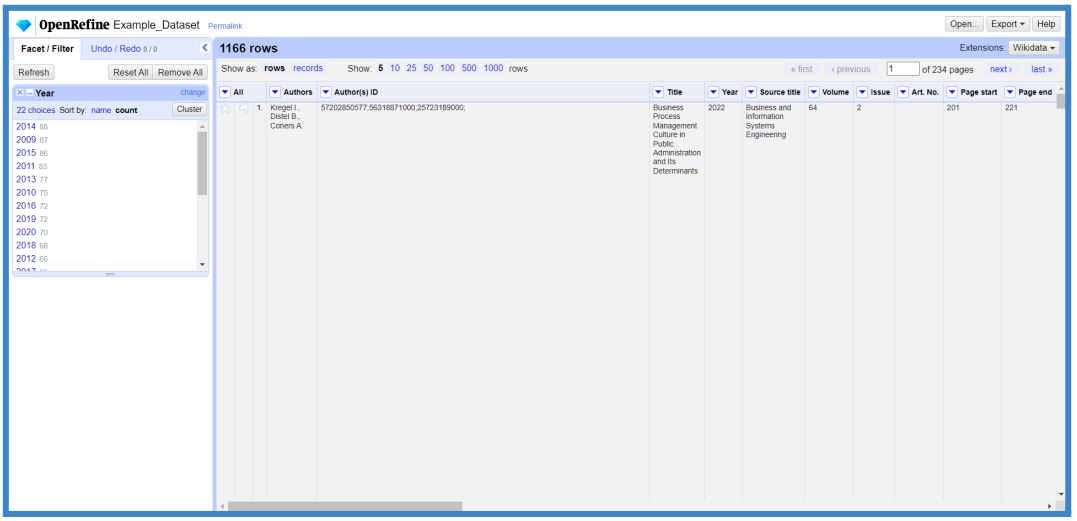


Figure 14. OpenRefine data exploration feature

3.5. Iteration: Database revision

Upon the software selection, the database selection was tackled again, to ensure that the database export options aligned with the software import options. As mentioned earlier (see Chapter 3.4.1) VOSviewer supports certain bibliographic databases and for each of them, the manual specifies the required format of the input file (Van Eck & Waltman, 2022). Subsequently, these databases were examined with regard to their limitations, in terms of the exported file formats and the amount of data, which can be downloaded at a time. (Dimensions Release Notes, 2020) (Lens Export Results, 2021) (PubMed User Guide, 2022) (Upgrades to Scopus: Export Limits Increased to 20,000, 2013) (Web of Science: Limit on Exporting, 2022) The distinction between them is presented in [Table 12].

Database	Exported file format[3]	Download limitation[4]
Web of science	Plain Text Tab-delimited	1.000
Scopus	CSV	20.000
Dimensions	CSV	2.500
Lens	CSV	50.000
PubMed	PubMed	10.000

Table 12. VOSviewer supported bibliographic databases: export format and limitations

From the table occurs that, in the case of Web of Science and Dimensions, some manual work might be necessary for data export; to run exports in multiple batches and then join the files together. The other databases seem to require no manual intervention, due to their high export limits. Referring back to the VOSviewer manual (Van Eck & Waltman, 2022) it was additionally taken into account that VOSviewer offers the possibility to create a network based on data from multiple files, thus joining files together would not pose a challenge. However, VOSviewer has another important limitation; the files must all be from the same source. Combining data from different databases is not possible. Thus it was elucidated that only one bibliographic database had to be selected.

The outcomes of the preprocessing procedure primarily suggest that the number of publications on “business process management” justify the deployment of Bibliometric Analysis. The

[3] Required by VOSviewer

[4] Measured in documents

VOSviewer-compatible databases meet the requirements related to the export format of the dataset and their export limits guarantee little to no required manual work. Among them, Scopus was put on the test and a high amount of results were obtained. Thus, the potential of the rest of the databases in this regard is yet to be discovered in order to choose the most suitable one for the analysis. The experimentations with the search query indicated that the exclusion of the "BPM" term from the search query grants more relevant results while the withdrawal of the "document type" limitation has no impact on the results. Lastly, the VOSviewer and OpenRefine software were found to be user-friendly and had complementary features thus were chosen to be used for the analysis of the results.

At the end of the preprocessing stage, a high level of awareness about the potential of the field, the bibliometric methods, and the available tools and software has been achieved. Hence the subsequent Bibliometric Analysis process can be deployed with a high level of confidence.

4. Stage 2: Bibliometric Analysis process

In this chapter the Bibliometric Analysis is deployed according to the flowchart of (Donthu et al., 2021) on the basis of the aims and objectives of the current study. The goal is to provide a very detailed application of every step, accompanied by visual elements that will help with the comprehension of each procedure.

4.1. Step 1: Aim and scope definition

In the first step of the Bibliometric Analysis, the scope of the study was primarily considered. Even though the "business process management" research field seemed initially broad it was thought that the "management" aspect might be a limiting factor thus it was decided to exclude it and broaden the scope of the study towards "business processes".

With this in mind, the Bibliometric Analysis aims were set in accordance with the initial objectives of the thesis, in the form of questions

- Objective 2: To examine the degree to which the field is shaped by diverse contributions and multidimensional aspects
    - Who are the most influential authors in the "business processes" research until 2022 and which are the associations among them?
    - Which countries invest and contribute the most to the development of the "business processes" and which are the associations among them?
- Objective 3: To discover the trends in the field which can reflect the major areas of future development
    - Which are the main topics of research in the business processes field?
    - Do studies orient their work towards marrying "business processes" with new advances in industry and technology?

4.2. Step 2: Technique selection

In alignment with the aim of the Bibliometric Analysis, it was decided that for the study of the interactions between constituents, Science Mapping techniques were the most suitable. From them, the Co-authorship and Co-occurrence analysis were chosen to be deployed, since they were the ones that could fulfil the aim and answer the research questions. The other three methods were not taken into account, since they address the impact of publications and their thematic relations.

4.3. Step 3: Data collection

4.3.1. Design of search term

For the design of the search query “business process” was chosen to be the main keyword, in accordance with the scope of the study. The “BPM” and “business process management” terms were excluded from the pool of potential keywords, as a result of former considerations. In order to ensure a high degree of relevance of the retrieved studies, it was concluded that the keyword would be placed in quotation marks in order to obtain only exact matches of the keyword, without alternations. The relevance of the results was also assured by searching for the keyword only in the documents’ title, which would imply that this is the main topic that the paper addresses.

Besides the main keyword, the search was limited to the years 2000-2022, in order to examine only recent, rather than obsolete, scientific work. Ultimately, it was assumed that well-known bibliographic databases include publications authored in English, which is the official scientific language. However, the linguistic factor was included in the search query as an additional layer of security for acquiring papers that address and are accessible by the international scientific community.

4.3.2. Selection of database

After shaping the query, the search was conducted on the databases which were compatible with VOSviewer and were discussed in Chapter 3.4.1. The exact query formulation, as well as the results (i.e. documents retrieved) of each database, are shown in [Table 13].

From the table it is visible that Web of Science and PubMed provided significantly fewer results (1.837 and 74 documents respectively), thus they were excluded from the pool of options. The Dimensions database provided the highest amount of results (i.e. 29.223) however that probably emerged from the fact that the keyword was searched in both the title and the abstract, not solely in the title. Due to this limitation, the database was disregarded as well. Out of the last two candidate databases, Lens (with 23.028 results) and Scopus (with 10.642 results), the former did not offer a linguistic filter, and it could not be identified whether the database includes publications exclusively in English. Even though Lens provided a double amount of results compared to Scopus, it was considered proper to select Scopus as the database for the data

retrieval, since it could run the whole query. In addition to that, the number of the results was considerably high and was considered representative of the research field.

Database	Search query	Results
Web of science	TI=("business process") AND PY=(2000-2022) AND LA=(English)	1.837
Scopus	TITLE ("business process") AND PUBYEAR > 1999 AND (LIMIT-TO (LANGUAGE , "English"))	10.642
Dimensions	Title and abstract: "business process" Filter: Publication year: 2000 - 2022	29.223
Lens	Title: ("business process") Filter: Year published = (2000 - 2022)	23.028
PubMed	"business process"[Title] Filter: Year: 2000 - 2022	74

Table 13. Search query in different databases

4.3.3. Data retrieval

In total, 10.642 documents were retrieved from Scopus on 06.06.2022 [Figure 15]. The "CSV export" option was selected and the "Export document settings" had to be set [Figure 16]. In the experimental preprocessing phase of Bibliometric Analysis, VOSviewer displayed errors with the imported CSV sample files. With more in-depth research, it was discovered that the errors had to do with the "Correspondence address" field, under "Bibliographical information" due to the fact that addresses all over the world can contain special characters that were not recognized. For this reason, this field was unchecked. The "Funding details" were also excluded from the downloaded information. From the "Other Information" section, only the "Include references" field was selected.

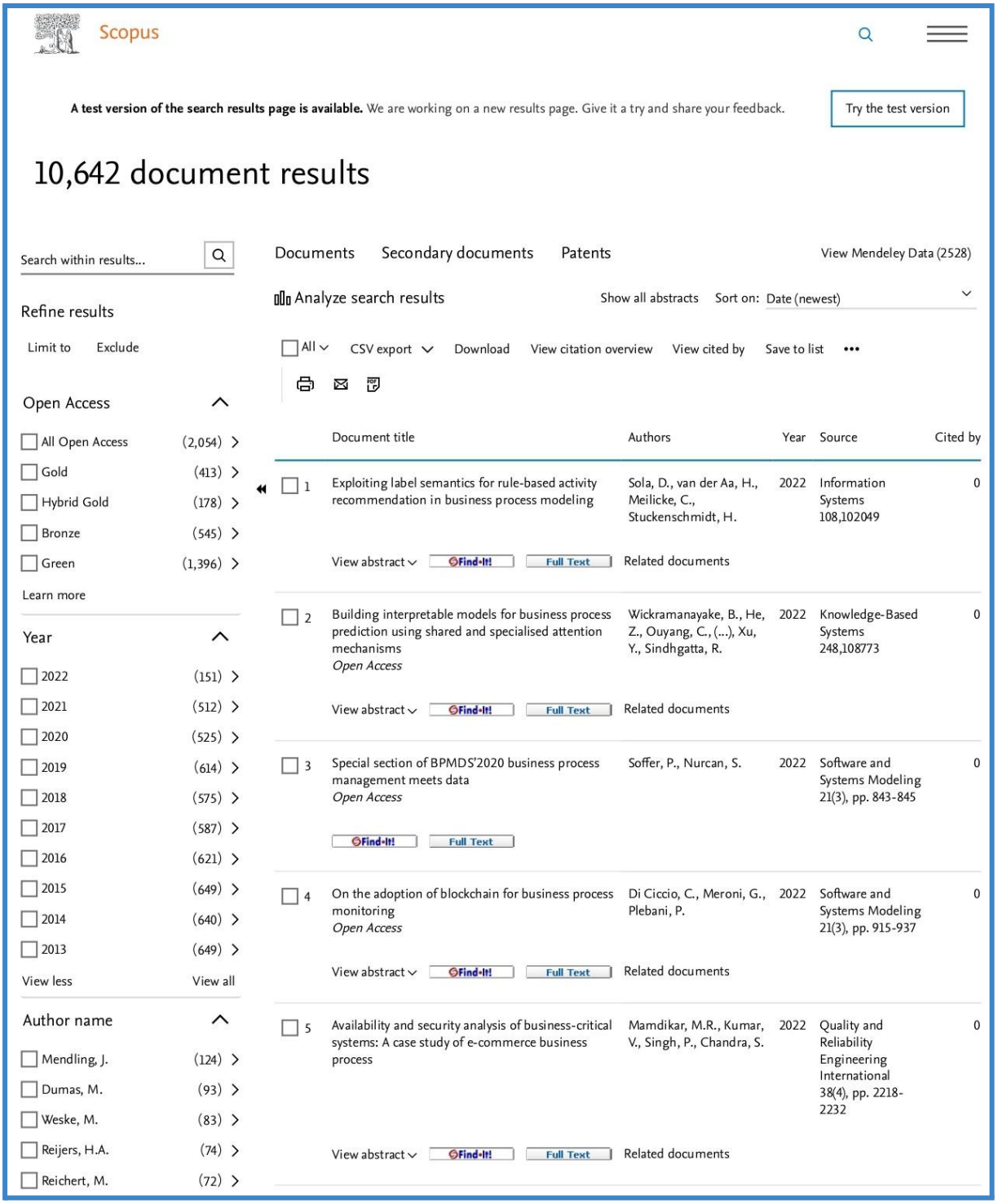


Figure 15. Scopus results

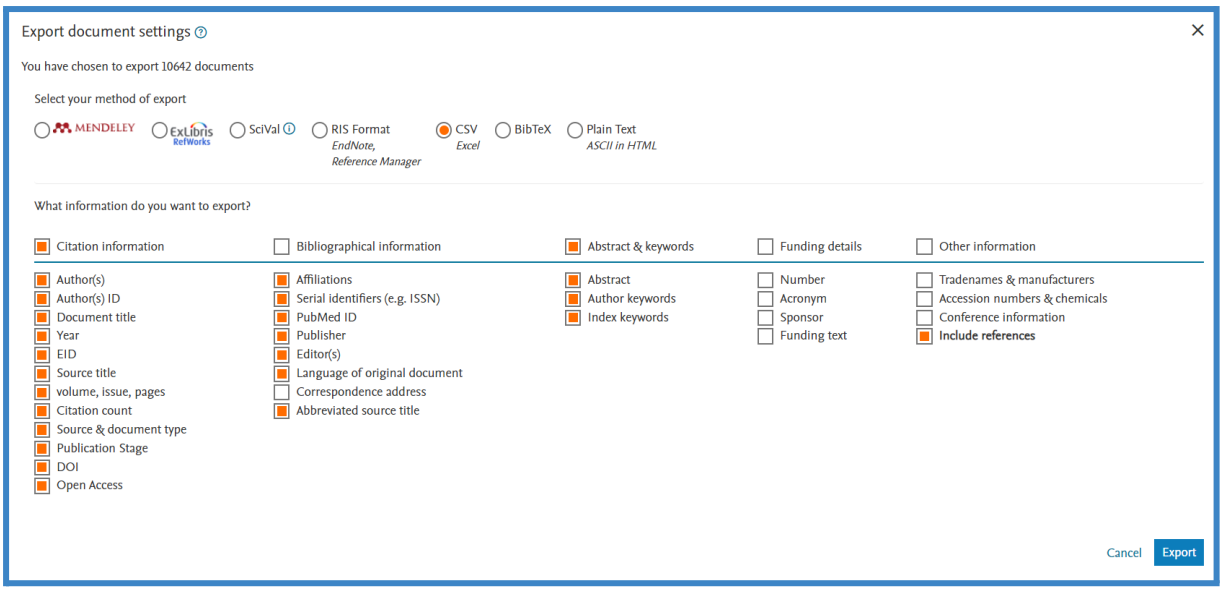


Figure 16. Scopus: export document settings

In the last screen of the settings [Figure 17], it was realized that the export limitation of 20.000 documents (Upgrades to Scopus: Export Limits Increased to 20,000, 2013) comes with the price of downloading only the citation information, meaning

- Author(s)
- Document title
- Year
- Source title
- Volume, issue, pages
- Citation count
- Source and document type

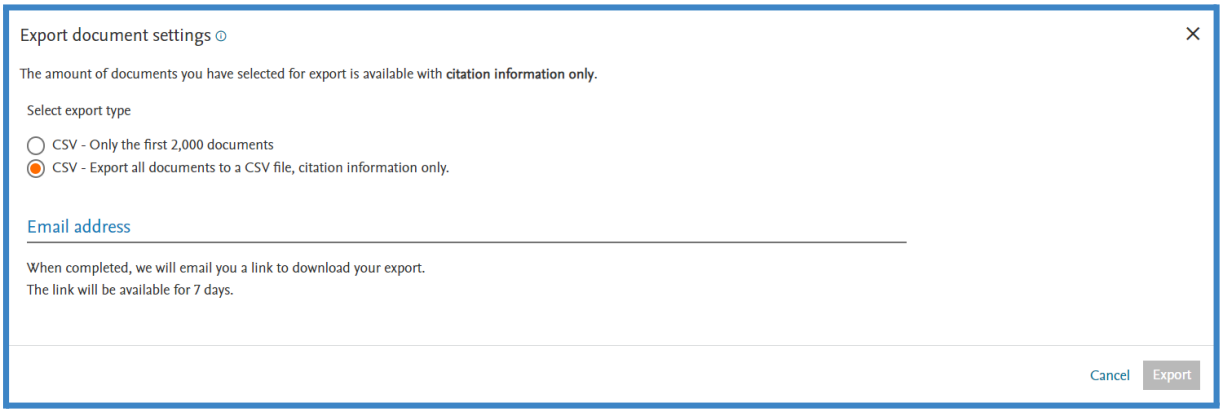


Figure 17. Scopus: export type settings

For this reason, at first, the 10.642 documents were downloaded as a batch, but the process was reiterated and the documents were downloaded again, in smaller batches, to retain all the necessary information.

The challenge in the latter case was that the option for downloading a smaller batch was bounded to "Only the first 2.000 documents". One way around it was to set the "Display of results per page" to the maximum (i.e. 200) and export batches of 200 documents. However, this way required a lot of manual labour. The ideal solution found was to use the filter "year" and create batches based on annual ranges with the limitation of a maximum of 2.000 documents per batch [Table 14].

Year range	Documents
2000-2007	1.816
2008-2010	1.876
2011-2012	1.427
2013-2015	1.938
2016-2018	1.783
2019-2022	1.802

Table 14. Scopus: export batches

4.3.4. Data cleansing

In the next step, the obtained data was cleansed. The six datasets were imported to OpenRefine one by one. In order to assure the validity of their content to some extent, the first operation performed was to exclude the documents, which had both the "Author" and "References" fields blank. If a document had entries, either in the "Author" or the "References" field, it was still considered valuable. More specifically, each CSV was imported in OpenRefine [Figure 18] and the default settings of the parse options were kept [Figure 19].

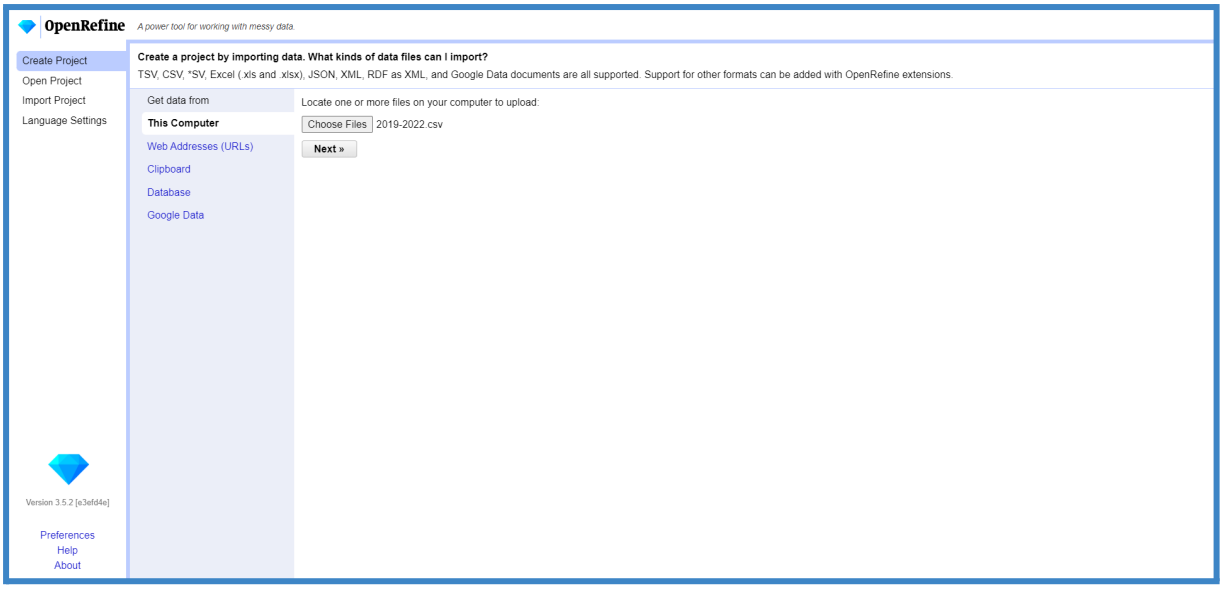


Figure 18. OpenRefine: data import

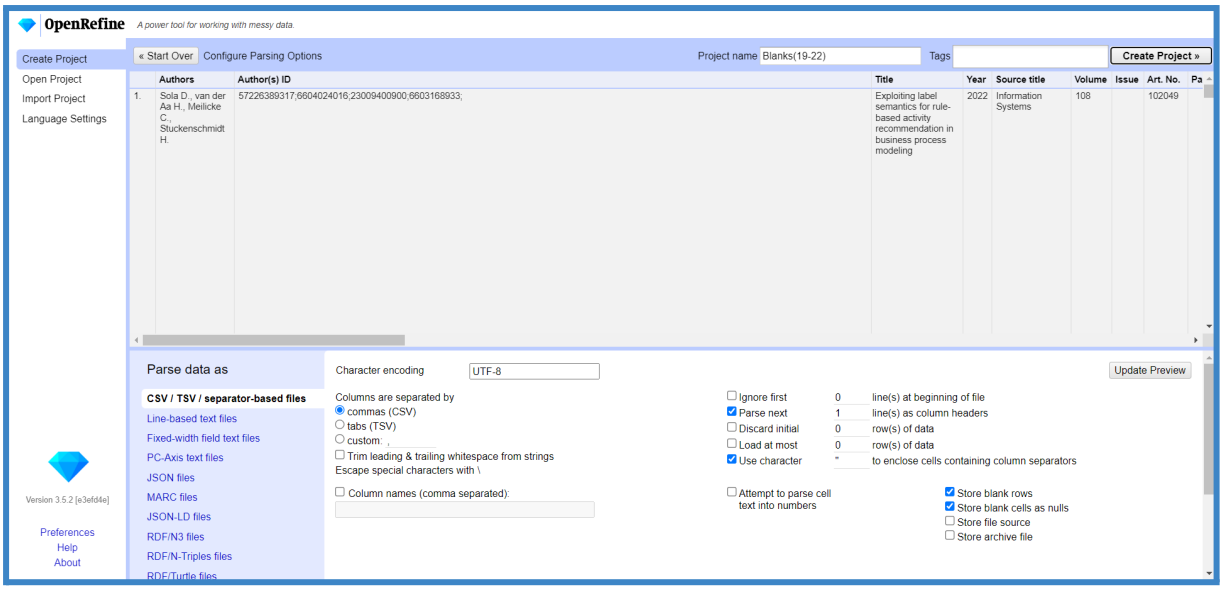


Figure 19. OpenRefine: configuration of parsing options

In the main interface [Figure 20] the first step was to explore the “Authors” records [Figure 21] with the “Text facet” option.

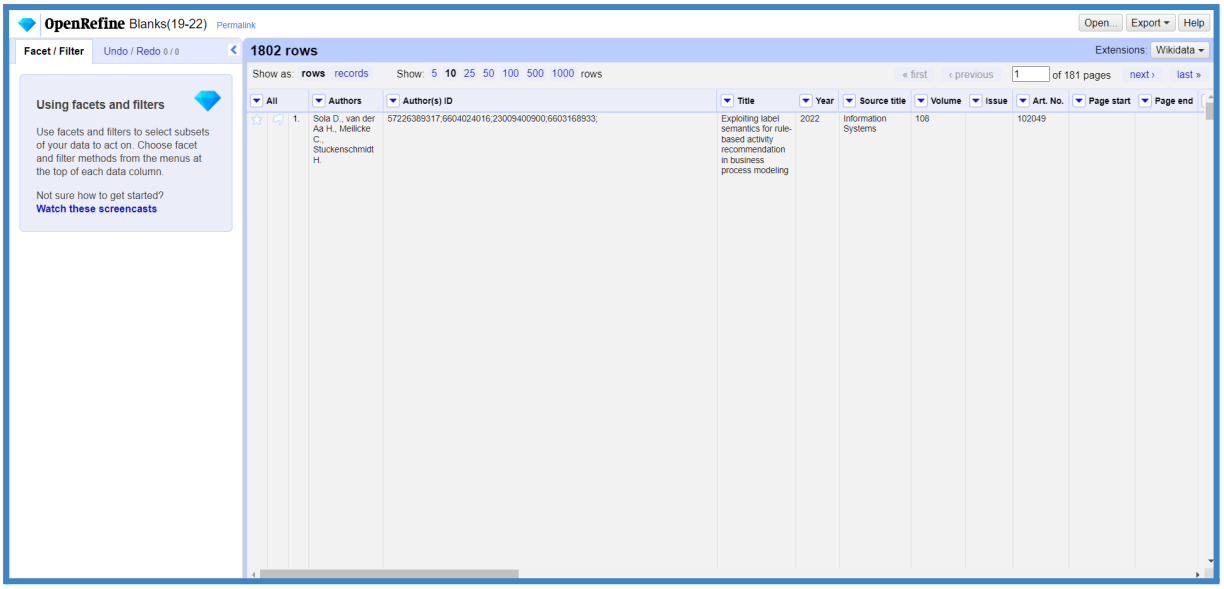


Figure 20. OpenRefine: documents overview interface

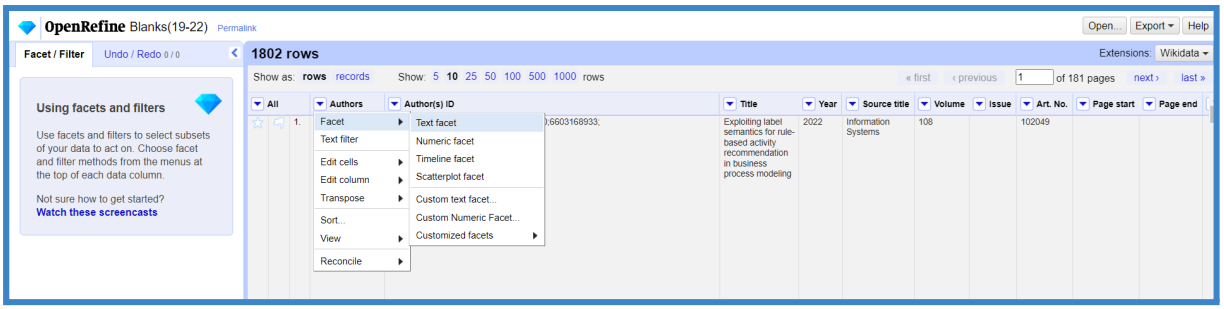


Figure 21. OpenRefine: "Authors" records exploration

In the "Authors" facet window [Figure 22] the documents with blank entries were spotted [Figure 23]. Afterwards, a "Text facet" was created for the "References" of the pre-selected documents [Figure 24].

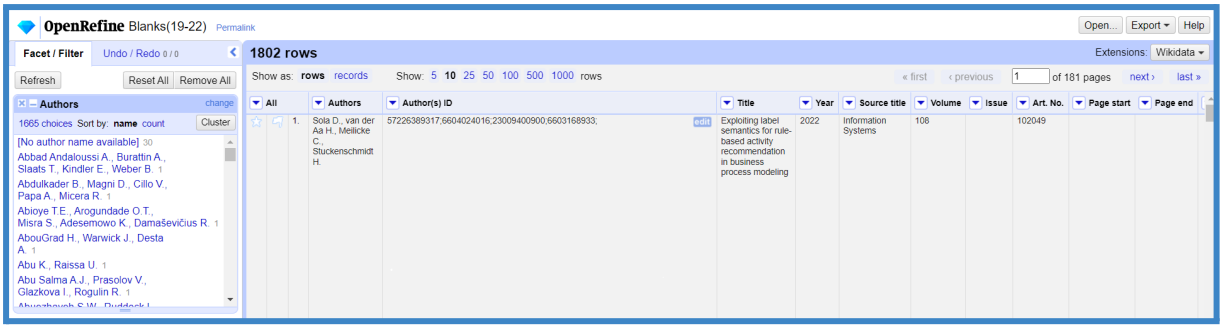


Figure 22. OpenRefine: "Authors" facet

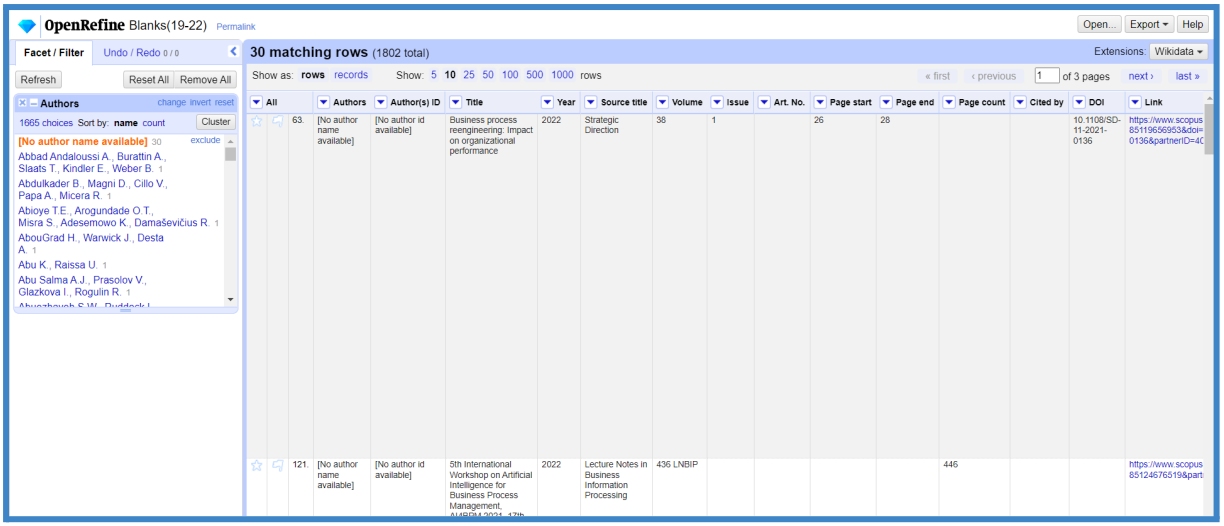


Figure 23. OpenRefine: Blank "Authors" records retrieval

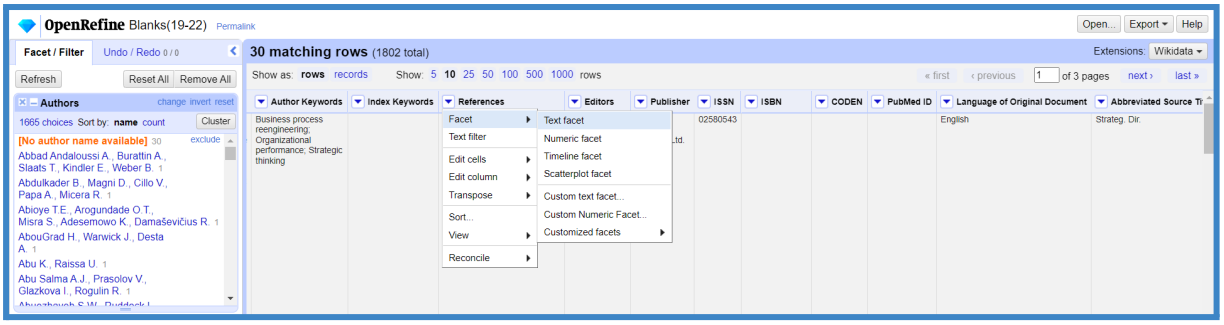


Figure 24. OpenRefine: "References" records exploration

Ultimately, the documents which had both fields empty [Figure 25] were removed from the dataset with the "Remove matching rows" option [Figure 26]. The refined dataset was exported in CSV format.

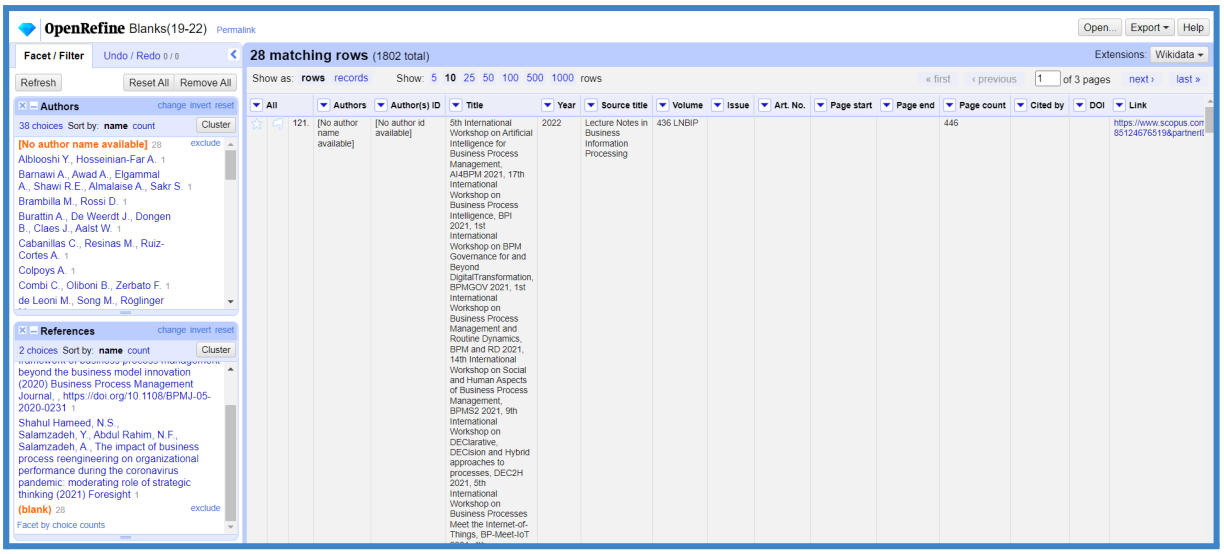


Figure 25. OpenRefine: Retrieval of documents with blank "Authors" and "References" records

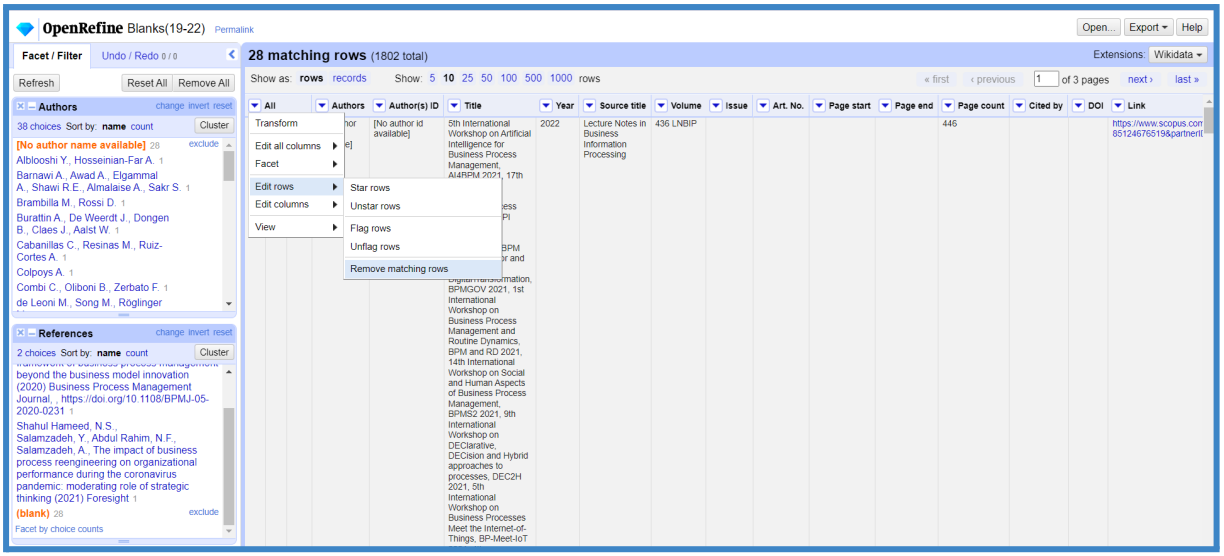


Figure 26. OpenRefine: Removal of documents with blank "Authors" and "References" records

Out of the total 10642 documents, 136 satisfied the condition of having both fields empty and they were removed from each dataset, resulting in 10506 documents [Table 15].

Document	Rows	Removed	Final rows
2000-2007	1.816	28	1.788
2008-2010	1.876	25	1.851
2011-2012	1.427	19	1.408
2013-2015	1.938	21	1.917
2016-2018	1.783	15	1.768
2019-2022	1.802	28	1.774
SUMMARY	**10.642**	**136**	**10.506**

Table 15. OpenRefine: Removal of rows with both the "Authors" and "References" fields empty

The next step of data manipulation addressed the "Author keywords" field. More specifically, the aim was to spot keywords with similar morphology and transform them into a unified form, to perform the subsequent Co-occurrence analysis.

The exported CSV files of the previous step were used as input, however, this time, altogether [Figure 27], in order to take into consideration the morphology of all the existing keywords. For the parsing options, the default settings were preserved again. By default, the "Author keywords" of each document were grouped within a cell and were separated by a semicolon. The "Split multi-valued cells" option was initially used [Figure 28] for splitting each group of keywords into distinct keywords based on the placement of the semicolon (semicolon and space, more

specifically) between them [Figure 29]. Ultimately distinct cells for each keyword were created [Figure 30]. The "Text facet" option was used afterwards to gain an overview of the existing keywords and their frequencies [Figure 31].

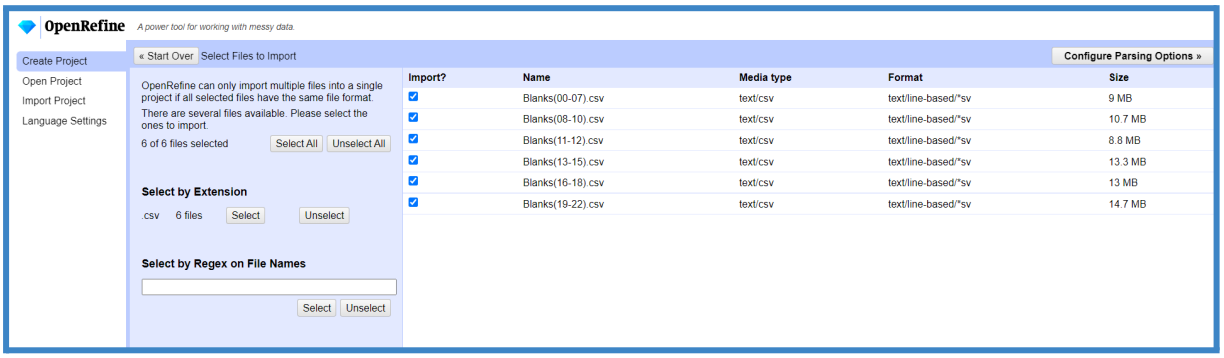


Figure 27. OpenRefine: Import of 6 CSV files

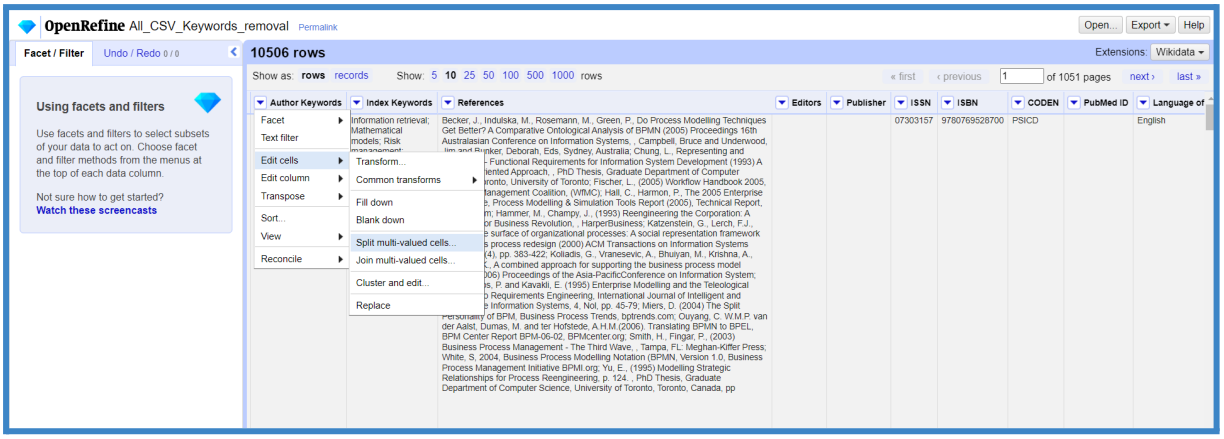


Figure 28. OpenRefine: Cell split option

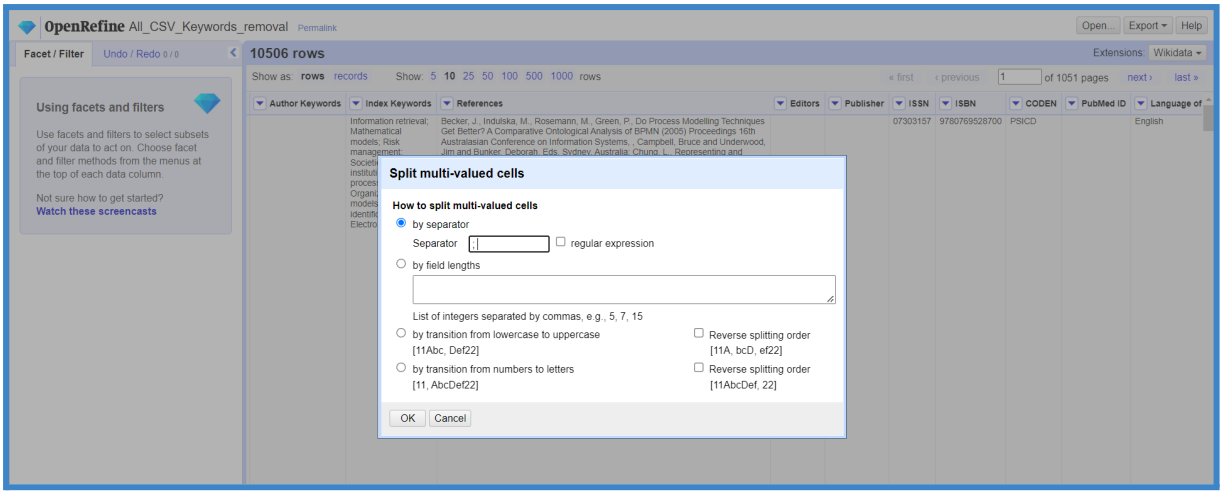
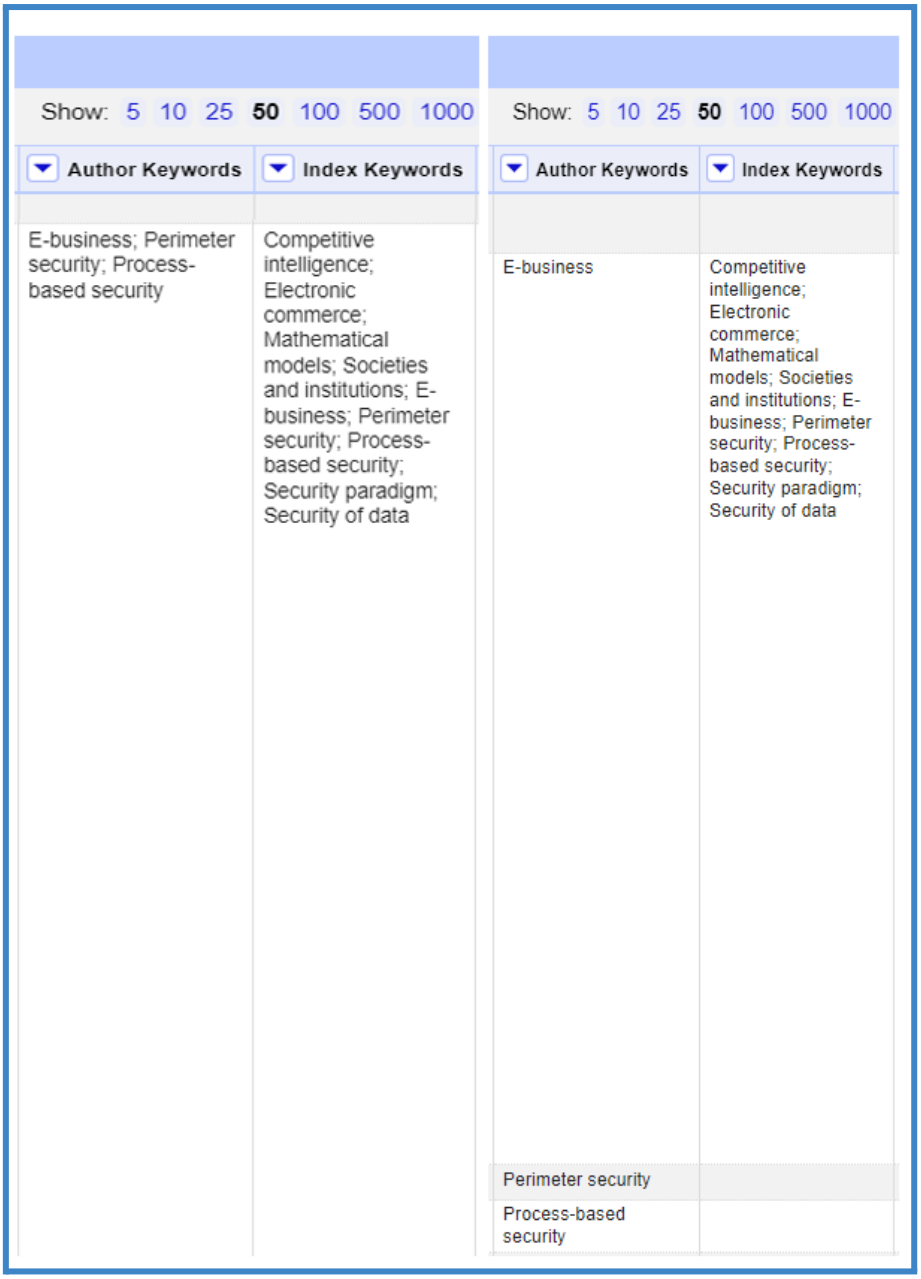


Figure 29. OpenRefine: Cell split condition



Show: 5 10 25 **50** 100 500 1000

Author Keywords	Index Keywords
E-business; Perimeter security; Process-based security	Competitive intelligence; Electronic commerce; Mathematical models; Societies and institutions; E-business; Perimeter security; Process-based security; Security paradigm; Security of data

Show: 5 10 25 **50** 100 500 1000

Author Keywords	Index Keywords
E-business	Competitive intelligence; Electronic commerce; Mathematical models; Societies and institutions; E-business; Perimeter security; Process-based security; Security paradigm; Security of data
Perimeter security	
Process-based security	

Figure 30. OpenRefine: "Author Keywords" cells before and after split

The descending sorting of the keywords by count revealed keywords that had to be merged, like "Business process", "Business Process" and "Business processes" or "Business process modeling" and "Business process modelling". OpenRefine has various clustering algorithms (OpenRefine User Manual, 2022) which cluster the keywords and spot keywords with similar morphology within each cluster. Then the algorithms propose a new morphology called "New cell value"; one word that will be used to replace all the similar instances of the term. The clustering process starts with the option "Cluster and edit" [Figure 32].

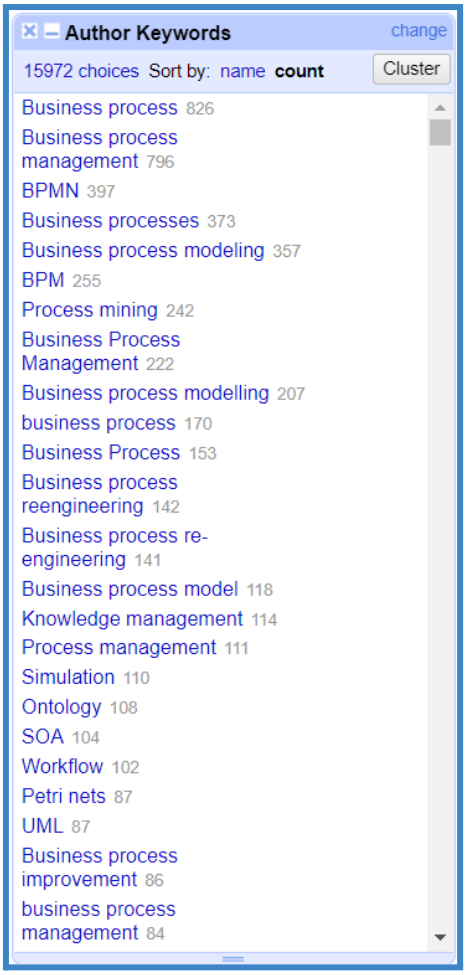


Figure 31. OpenRefine: Text facet of "Author Keywords" before clustering

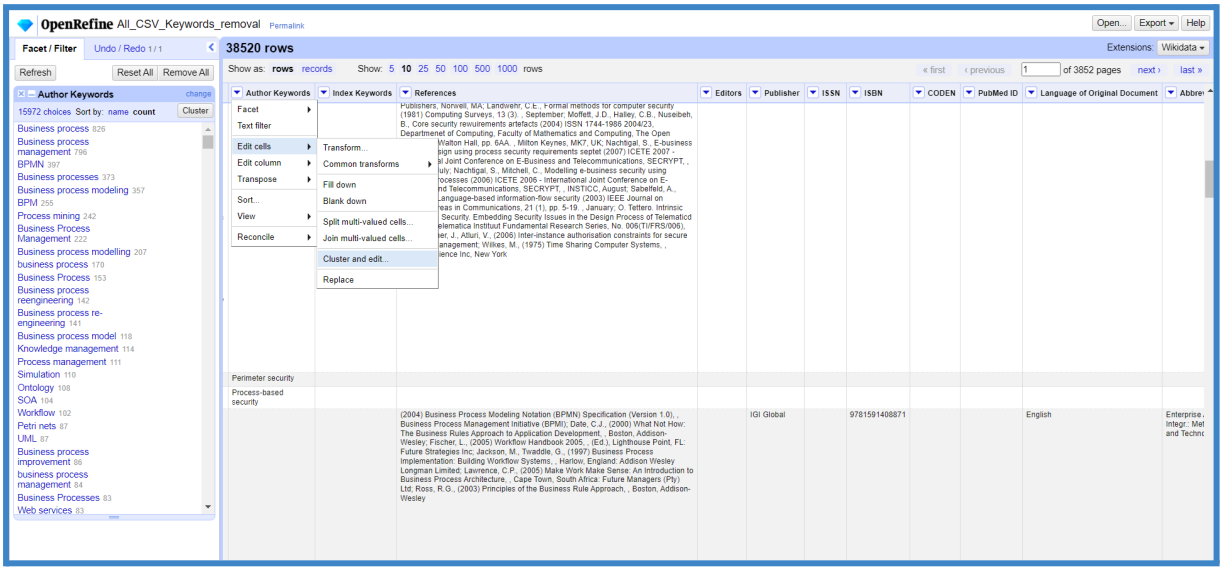


Figure 32. OpenRefine: Automatic clustering option

The default clustering method is “key collision” with the keying function “fingerprint” was initially performed. Key collision is very fast and can process millions of cells in seconds while fingerprinting performs the following operations (OpenRefine User Manual, 2022)

- fixes whitespace into single spaces
- puts all uppercase letters into lowercase
- discards punctuation
- removes diacritics (e.g. accents) from characters
- splits up all strings (words) and sorts them alphabetically (so “Zhenyi, Wang” becomes “wang zhenyi”)

After one iteration [Figure 33], the algorithm did not produce more clusters.

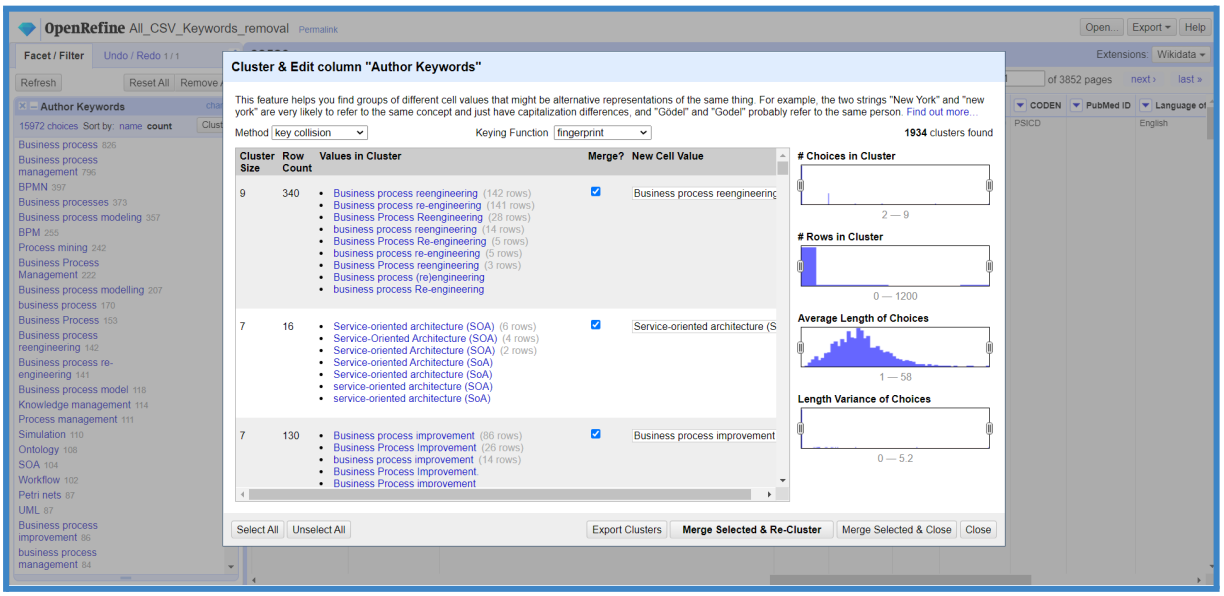


Figure 33. OpenRefine: First iteration of clustering in with method “key collision”

Then, the second default method, “nearest neighbour” using the “Levenshtein distance” algorithm with “Radius”=1.0 and “Block Chars”=6, was deployed. The latter method unfolded additional clusters and two more iterations were performed [Figure 34] [Figure 35].

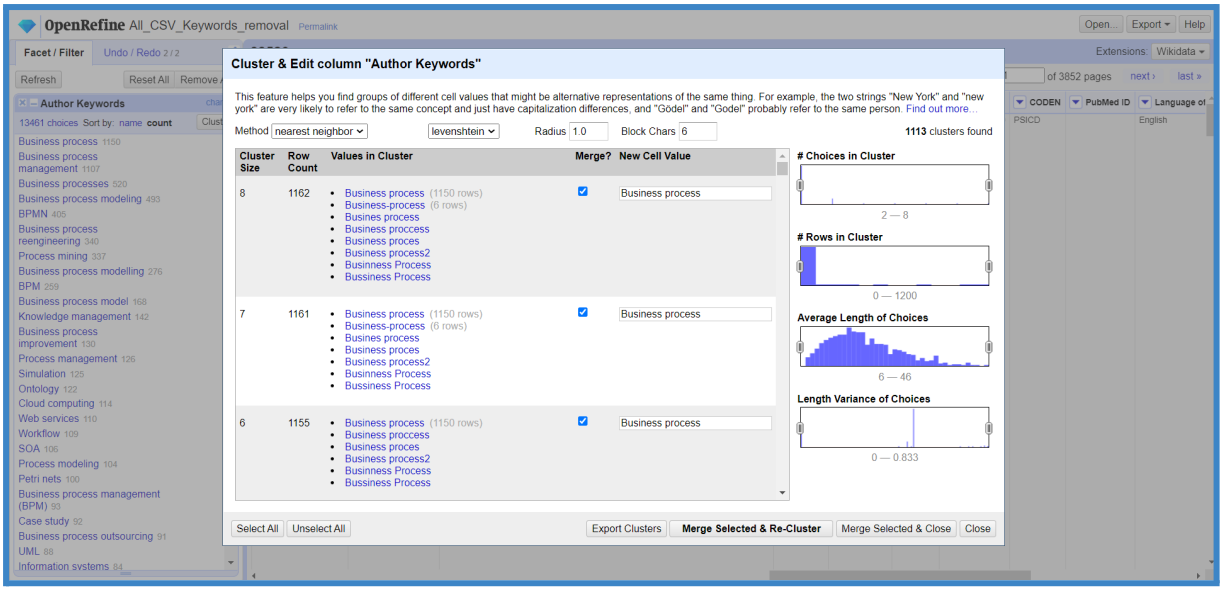


Figure 34. OpenRefine: First iteration of clustering with method "nearest neighbour"

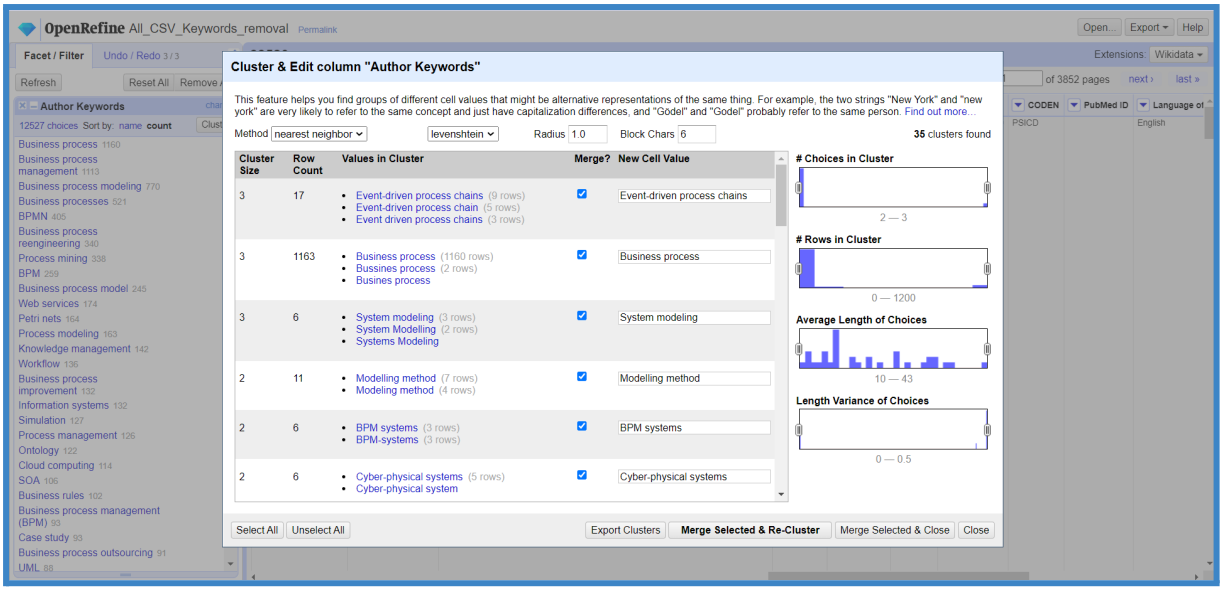


Figure 35. OpenRefine: Second iteration of clustering with method "nearest neighbour"

The automatic clustering was considered fairly accurate, however, when the keywords were enlisted again with the "Text facet" option [Figure 36], it was noticed that there were still instances that had not been clustered (e.g. "Business process" and "Business processes"), hence manual clustering was imperative.

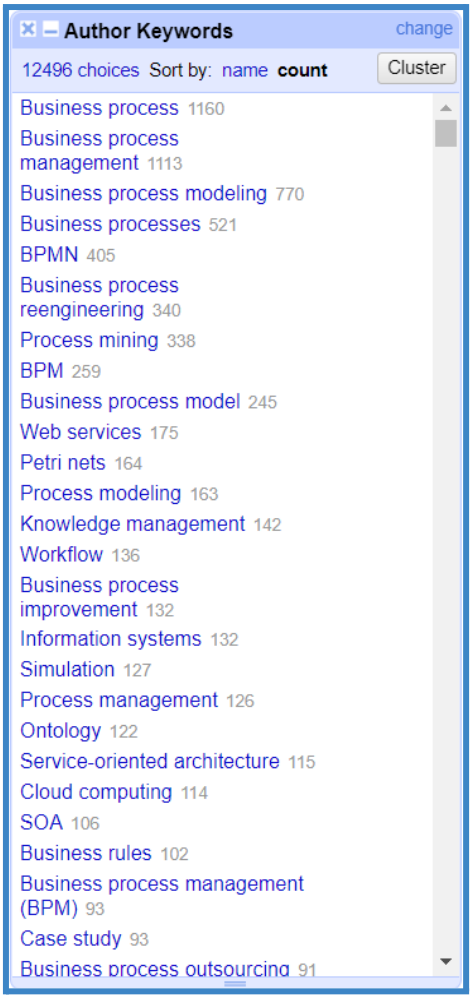


Figure 36. OpenRefine: OpenRefine: Text facet of "Author Keywords" after automatic clustering

The main modifications that were performed manually are shown in [Table 16]. The titles indicate the "New cell value" and the list below them comprises the terms with similar morphology that were replaced. It is worth noticing that the terms "Business process model" and "Business process modeling" were not merged because it was assumed that the term "model" refers to models generally, while "modeling" refers to a certain phase of the business process management lifecycle. Additionally, it was observed that "BPM" stood for either business process management, modelling, or mining while "BPR" stood for either business process re-engineering, redesign or review. Since these abbreviations had no explicit reference (e.g. Business Process Management (BPM)) the automatic algorithm clustered all the instances within the same cluster. [Figure 37] presents the keywords before clustering, after the automated clustering and after the manual intervention.

Business process	
Business processes	521
Business processes (BPs)	5
Business process reengineering	
Business process reengineering (BPR)	39
Reengineering	33
Process reengineering	16
Business Processes Re-engineering	3
Business re-engineering	3
Business process modeling	
Process modeling	163
Modelling	73
Modeling	69
Business processes modeling	17
BP modeling	11
Business process modeling framework	2
Business process modeling goal	2
Business process modeling techniques	2
Business process modelling (BPM)	13
Business process modelling tools	4
Business processes modeling (BPM)	2
Business process model	
Process model	63
Business process model abstraction	8
Business process model reuse	6
Business process model (BPM)	4
Business process model cost extension	4
Business process model comprehension	3
Business process model repository	3
Business process model standardization	2
Business process model design	2
Business process model discovery	2

Business process management	
Business process management (BPM)	93
Business process management system	58
Business processes management	11
Business processes management (BPM)	1
Process management	126
Business Process Modeling Notation	
BPMN	405
Business process modeling notation (BPMN)	11
Business process modeling notation	11
BPMN extension	10
BPMN Model	8
BPMN Process Model	4
Business processing modeling notation	1
Business process modeling notation diagrams	1
Business Process Model and Notation (BPMN)	23
Business process model and notation	20
transformation of Business Process Model and Notation	1
BPM Notation	6
Business process model notation	2
Business Process Model Notation, BPMN	2
Business Process Model Notations (BPMN)	1
Business Process Modeling and Notation	2
Business process improvement	
Process improvement	66
Business improvement	4
Business process improvement (BPI)	3
Business processes improvement	3
Process model improvement	2
(bp) Improvement	1

Table 16. OpenRefine: terms that were manually incorporated into bigger clusters

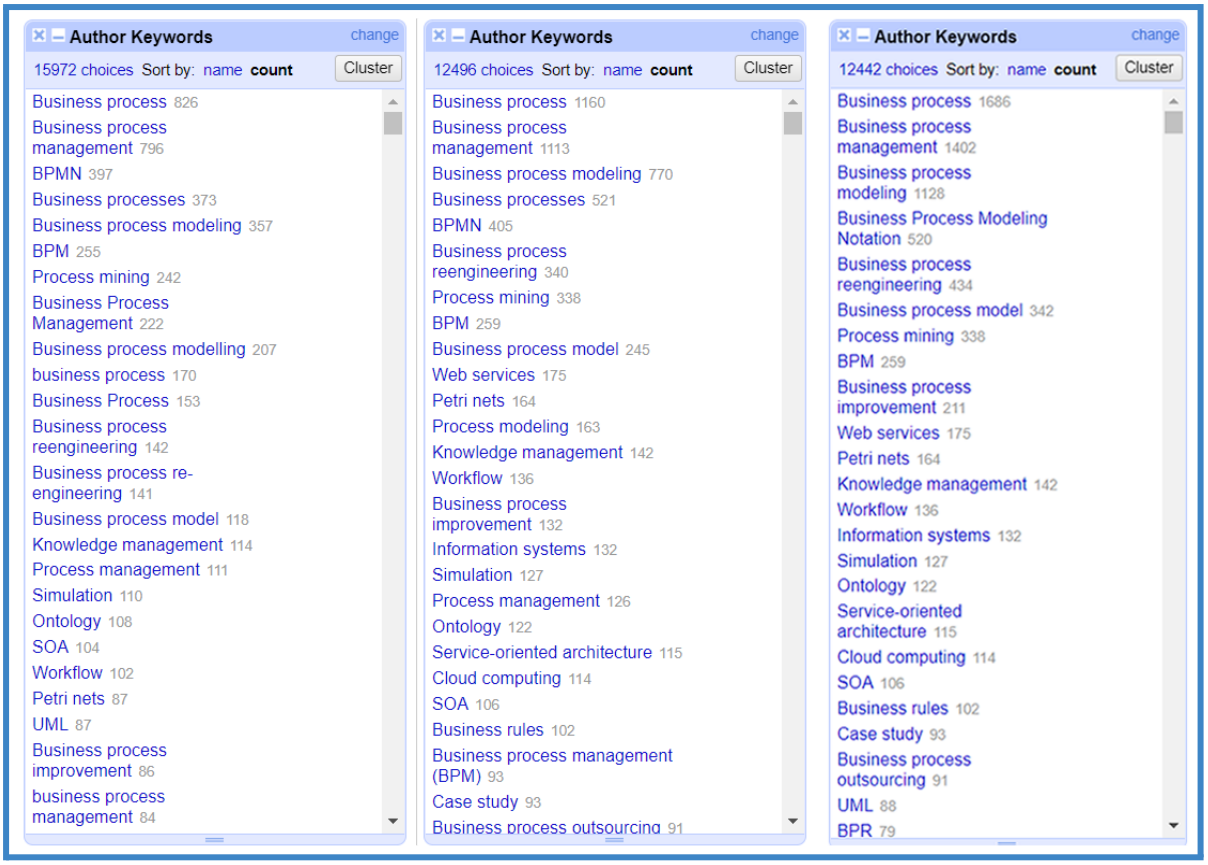


Figure 37. OpenRefine: Author keywords clusters before clustering (left), after automated clustering (middle), and after manual adaptations (right)

Upon clustering, the option "Join multi-valued cells" was selected [Figure 38] and the split cells were joined back together (i.e. the keywords of each document were grouped back together). All the steps of the clustering process were saved under the Undo/Redo tab and the operation history [Figure 39] was copied and saved to a text file. Eventually, the 6 imported datasets were exported as one file, in CSV format [Figure 40][5].

[5] *A realization that was made later was that VOSviewer cannot run the Co-authorship analysis (with "Authors" as the unit of analysis) with the previous, merged CSV file as input [Figure 41]. However, VOSviewer supports the import of multiple files. For this reason, the same clustering steps were applied again to each of the 6 CSV files separately. More specifically, the "Apply" functionality of OpenRefine was exploited (in the Undo/Redo tab) and the formerly exported operation history was applied to each CSV file separately. Later on, the files were imported into VOSviewer and the Co-authorship analysis was performed.*

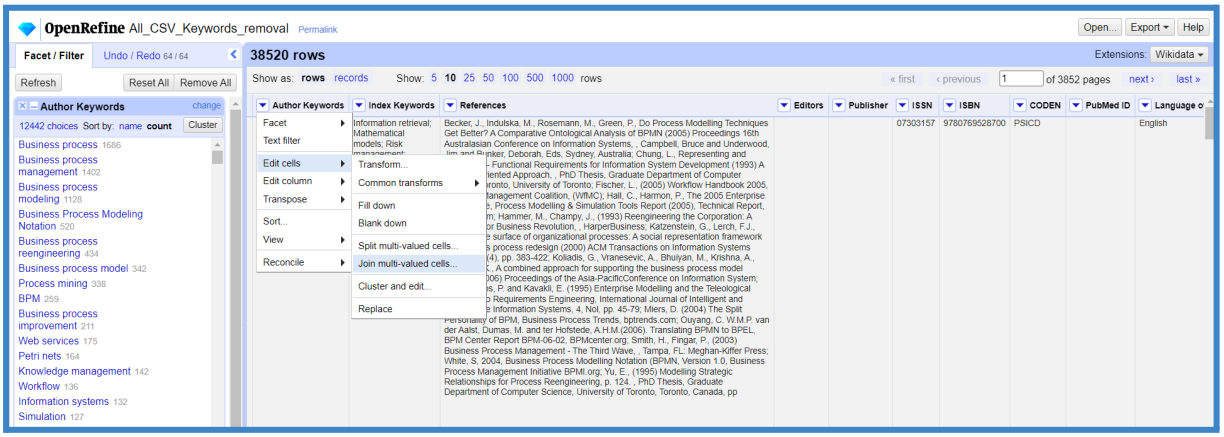


Figure 38. OpenRefine: Join of split cells

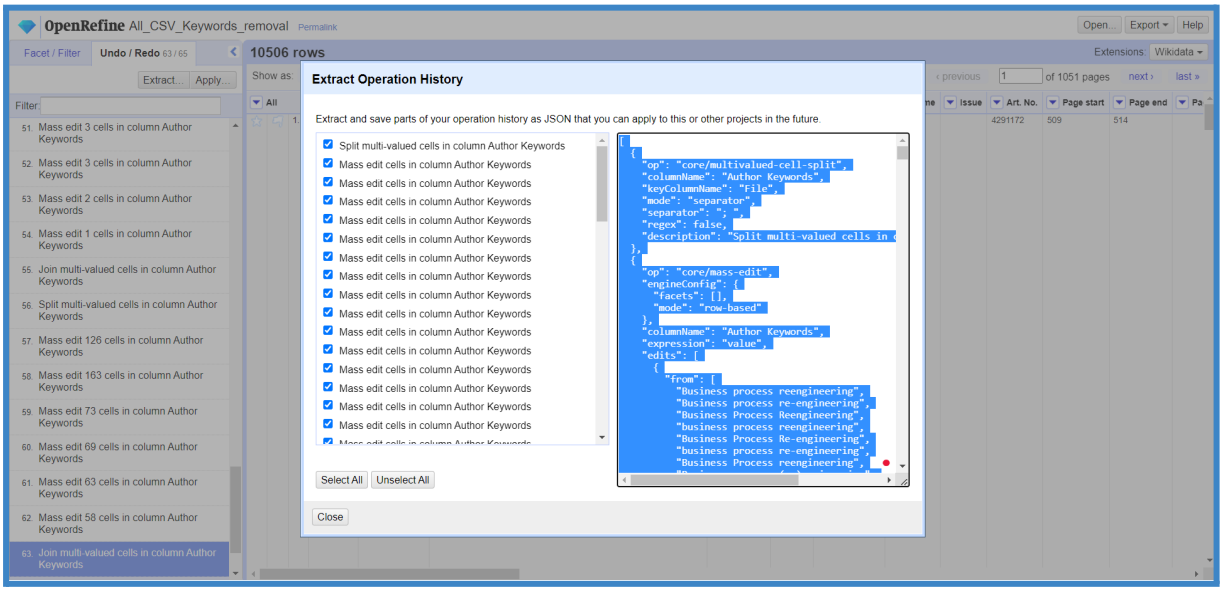


Figure 39. OpenRefine: Extraction of operation history

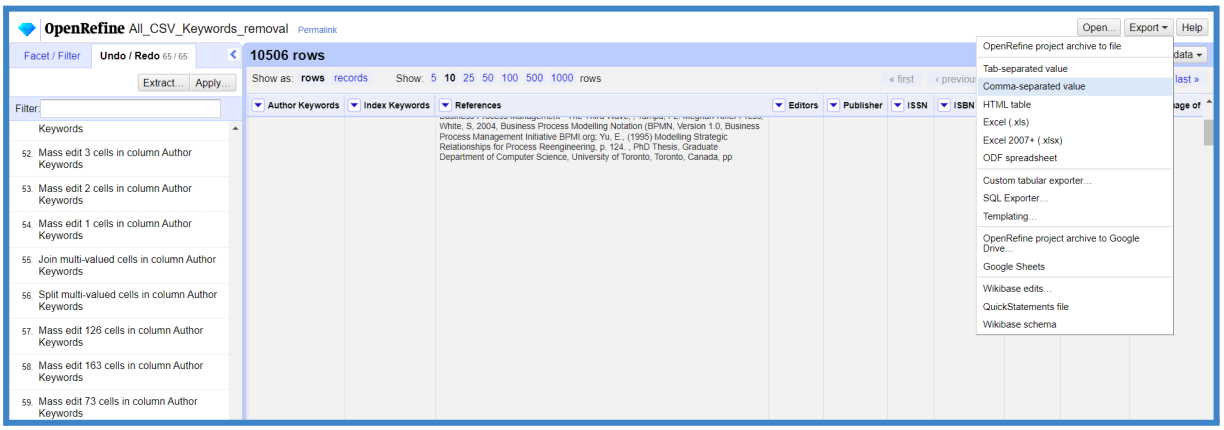


Figure 40. OpenRefine: Export in CSV

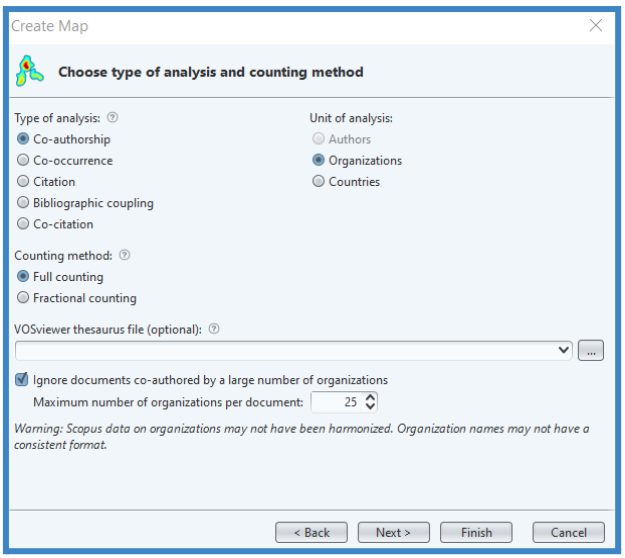


Figure 41. VOSviewer: deactivated Co-authorship option with merged CSV file as input

4.4. Step 4: Data analysis

After the dataset was cleansed, it was ready to be further analyzed. As described in Chapter 5.4.1. VOSviewer was selected as the software for the data analysis, mainly because of its compatibility with bibliographic database files, its interactive output, the various export options that it offers, and its user-friendly interface.

Besides that, VOSviewer has additional features: it is able to create networks automatically, without the necessity of additional software, it provides various approaches for each Science Mapping technique (e.g. Co-authorship with either author, country or institution as the unit of analysis), it offers considerable configuration options for each technique and it supports various types of data visualization. For a more profound understanding of these functionalities the basic terminology, the visualization options, and the data import process are described in more detail.

4.4.1. Terminology

In VOSviewer, a network is a set of items together with the links between them. **Items** are the objects of interest (e.i. publications, researchers, terms, etc) and a map normally includes only one type of item. Between any pair of items, there can be a unique **link**, which represents the connection or relation between these items.

Each link between items has a **strength** (i.e. a positive numerical value). The strength of a link may for example indicate the number of cited references two publications have in common (Bibliographic coupling), the number of publications two researchers have co-authored

(Co-authorship analysis), or the number of publications in which two terms occur together (Co-occurrence analysis). The higher the strength, the stronger the link.

Items may be grouped into non-overlapping **clusters** meaning that an item may belong to only one cluster. However, there may be items that do not belong to any cluster. The clusters are labelled using cluster numbers.

Items may have various **attributes**. If items have been assigned to clusters, the cluster numbers are an example of an attribute. Two important attributes are the **weight** and the **score**, both represented by numerical values. The weight of an item indicates its importance, thus the higher the weight the higher the importance of an item. In the visualization of a map, items with a higher weight are shown more prominently than items with a lower weight. There are two standard weight attributes: the **Links** attribute, which indicates the number of links of an item with other items, and the **Total Link Strength (TLS)** attribute, which indicates the total strength of the links of an item with other items. A score attribute may indicate any other numerical property of items besides their importance, which is reflected in their weight (Van Eck & Waltman, 2022).

4.4.2. Visualization

The VOSviewer provides three types of visualizations:

- network visualization
- overlay visualization
- density visualization
    - item density
    - cluster density

In the **network visualization**, items are represented by their label and - by default - by a circle. The higher the weight of an item, the larger the label and the circle of the item. The **overlay visualization** is identical to the network visualization however the colour of the items is determined by their scores (the default colours range from blue (lowest score) to green to yellow (highest score)) [Figure 42]. The density visualization has two variants: item density and cluster density. In the **item density visualization**, items are represented by their label in a similar way as in the other two visualization types. Each point in the item density visualization has a colour that indicates the density of items at that point. By default, colours range from blue to green to yellow. Thus, the larger the number of items in the neighbourhood of a point and the higher the

weights of the neighbouring items, the closer the colour of the point is to yellow. The **cluster density visualization** is identical to the item density visualization except that the density of items is displayed separately for each cluster of items [Figure 43] (Van Eck & Waltman, 2022).

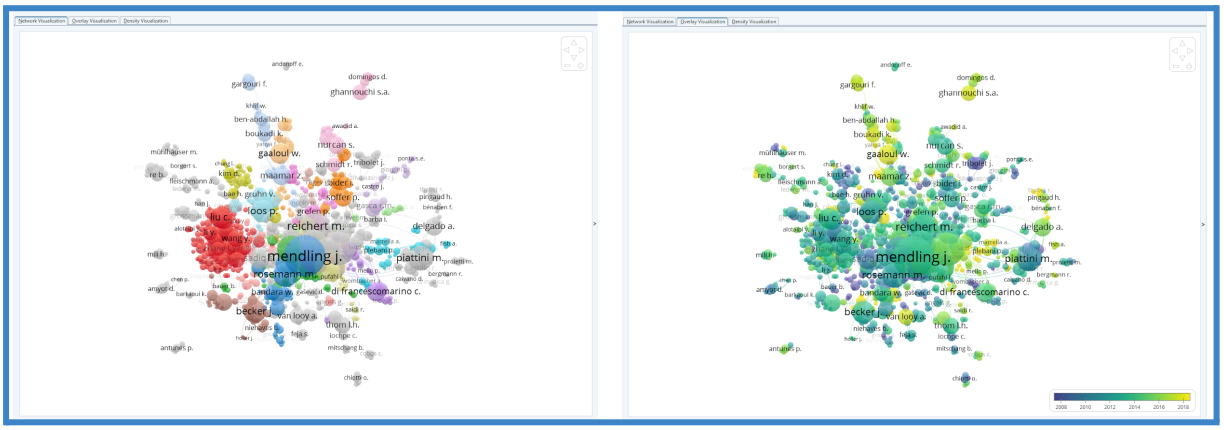


Figure 42. VOSviewer: Network visualization (left) and Overlay visualization (right)

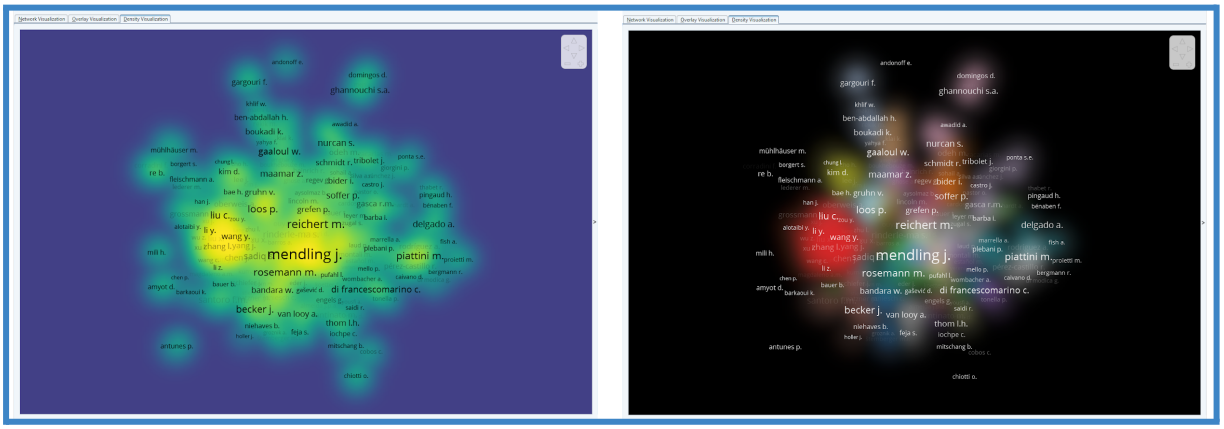


Figure 43. VOSviewer: item density visualization (left) and cluster density visualization (right)

4.4.3. Data import

In the main window of VOSviewer the menu on the left is called “action panel” while the menu on the right is called “options panel” [Figure 44] For the creation of a network out of data of a CSV file (or more) exported from Scopus, the option “Create” is initially selected from the “action panel”. Then the type of imported data is selected (for this study the “Create a map based on bibliographic data” option) [Figure 45]. Afterwards, the data source is selected (for this study the “Read data from bibliographic database files” option) [Figure 46]. Next, the dataset is imported into the tab “Scopus” [Figure 47] and the type of analysis is selected [Figure 48]. The items which can be visualized with VOSviewer for each type of analysis are shown in [Figure 49]. These are the main steps, which can lead to the creation of a network, with the possibility of customization of certain options.

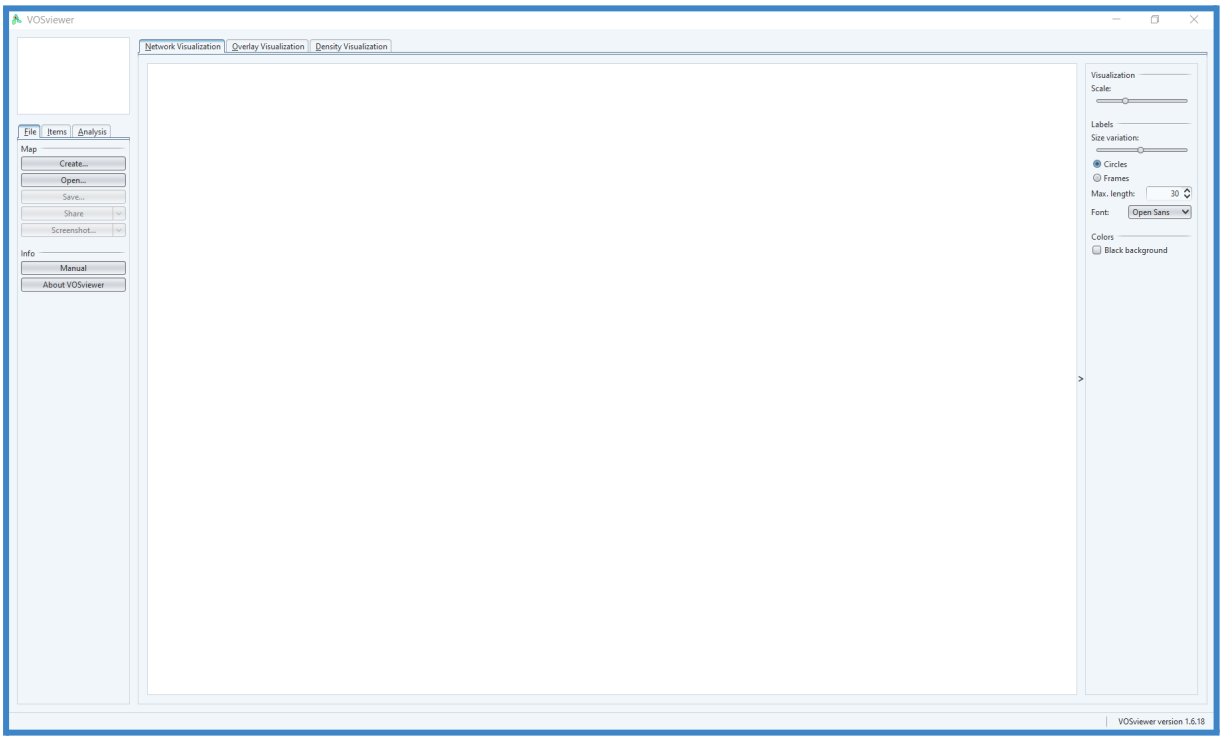


Figure 44. VOSviewer: Main window with action panel (left) and options panel (right)

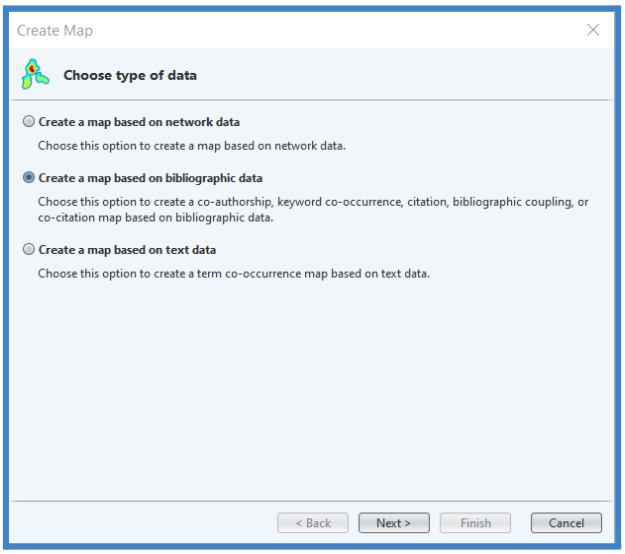


Figure 45. VOSviewer: Selection of type of imported data

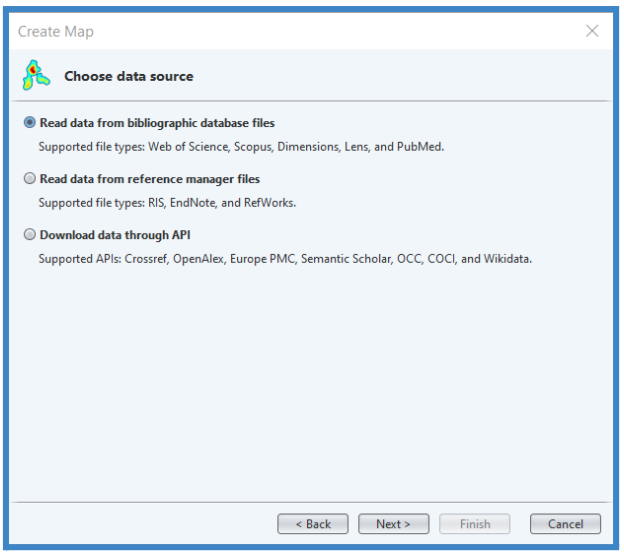


Figure 46. VOSviewer: Selection of data source

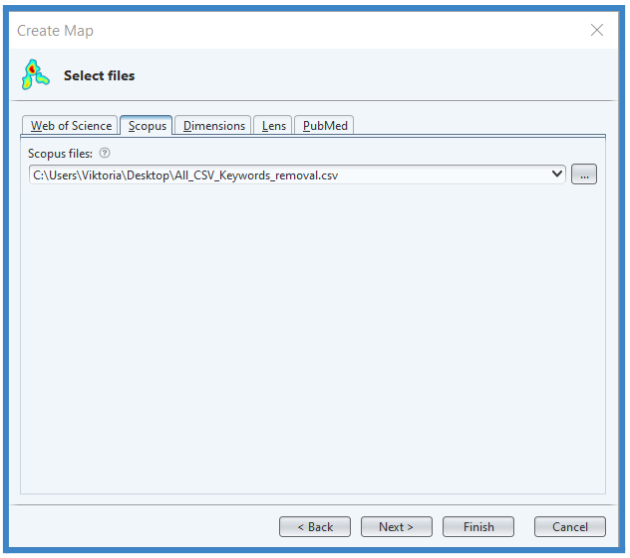


Figure 47. VOSviewer: Data import

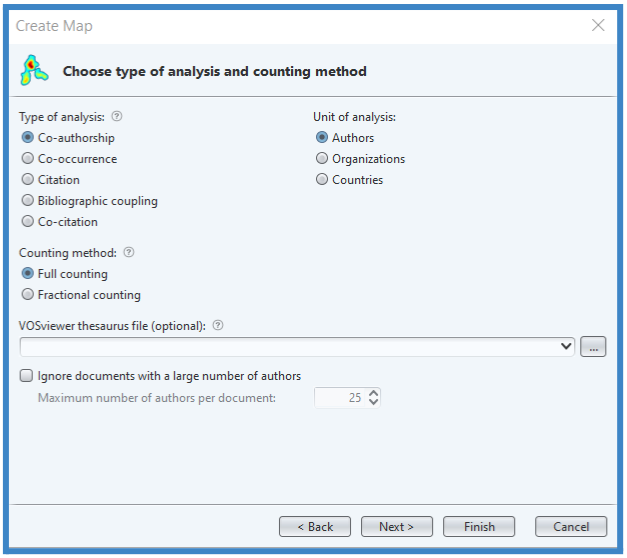


Figure 48. VOSviewer: Selection of type of analysis

Data	Files	Links	Items
Bibliographic data	Web of Science files	Co-authorship	Authors
	Scopus files		Organizations
	PubMed files		Countries
	RIS files	Co-occurrence	Keywords
	Crossref JSON files	Citation	Documents
			Sources
			Authors
			Organizations
			Countries
		Bibliographic coupling	Documents
			Sources
			Authors
			Organizations
			Countries
		Co-citation	Cited references
			Cited sources
			Cited authors

Figure 49. VOSviewer: Items that can be visualized in each type of analysis (Van Eck & Waltman, 2022)

At the end of the Bibliographic Analysis process, the scope of the study has been set and the query has been formulated accordingly. Based on the aim set, the bibliometric techniques have been chosen and the data have been retrieved and cleansed. Ultimately, the dataset is imported into the selected software and, upon possible adjustments of the options, the results can be obtained. The results for each of the selected Science Mapping techniques are analytically presented and discussed in the next chapter.

5. Results

In this chapter the application of the selected Bibliometric Analysis techniques (Co-authorship and Co-occurrence analysis) is thoroughly presented. The chapter aims to provide a comprehensive guide on the deployment of each technique with the use of the features of VOSviewer software and, ultimately, answer the research questions that were posed in Stage 2 of the Bibliometric Analysis process.

5.1. Co-authorship analysis (Authors)

The first type of analysis performed was the Co-authorship analysis, with "Authors", as the unit of analysis. As mentioned in Chapter 4.3.4, this type of analysis was not possible with the use of the unified CSV file, thus the 6 separate CSV files were used as input. The option to "ignore documents co-authored by a large number of authors" was disabled [Figure 50] and the fields "minimum number of documents of an author" and "minimum number of citations of an author", were left with their default value; 5 and 0 respectively [Figure 51]. As indicated by the software, out of the 15246 authors only 1161 met these thresholds and out of these, 841 were connected (the rest did not have any links) [Figure 52].

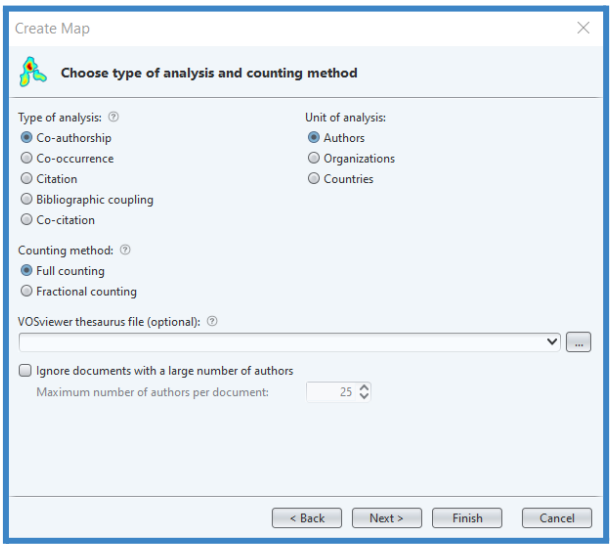


Figure 50. VOSviewer: window for selection of Co-authorship as the type of analysis with Authors as the unit of analysis

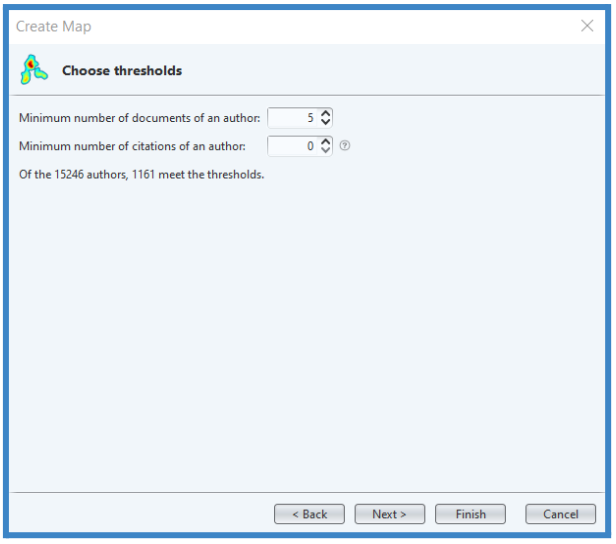


Figure 51. VOSviewer: Co-authorship (authors) window for selection of thresholds

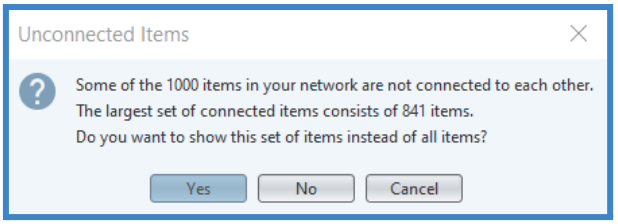


Figure 52. VOSviewer: Co-authorship (authors) warning message for connected items

The network was created based on the weight attribute "TLS" (set in the "options panel") [Figure 53]. The nodes represented the authors and their size was proportional to the TLS of each author (bigger size denoted higher TLS). As mentioned in Chapter 5.1 in Co-authorship analysis "Researchers are linked to each other based on the number of publications they have authored jointly". Consequently, in the produced network, the links between authors indicated the existence of co-authorship. Each link between two authors had a "**Link strength"** attribute, denoting the times that they have collaborated on publications (i.e. the total number of co-authored documents). When looking at each author individually, the unique number of authors that one had collaborated with was reflected in the "**Links"** attribute (a maximum number of links could be 840, implying that the author had collaborated with all the authors present in the map). The "**TLS"** attribute of an author, which was reflected on the node size, indicated the cumulative "**Link strength"** of all of its "**Links"**. Additionally, each author had "**documents**" and "**citations**" as additional weight attributes.

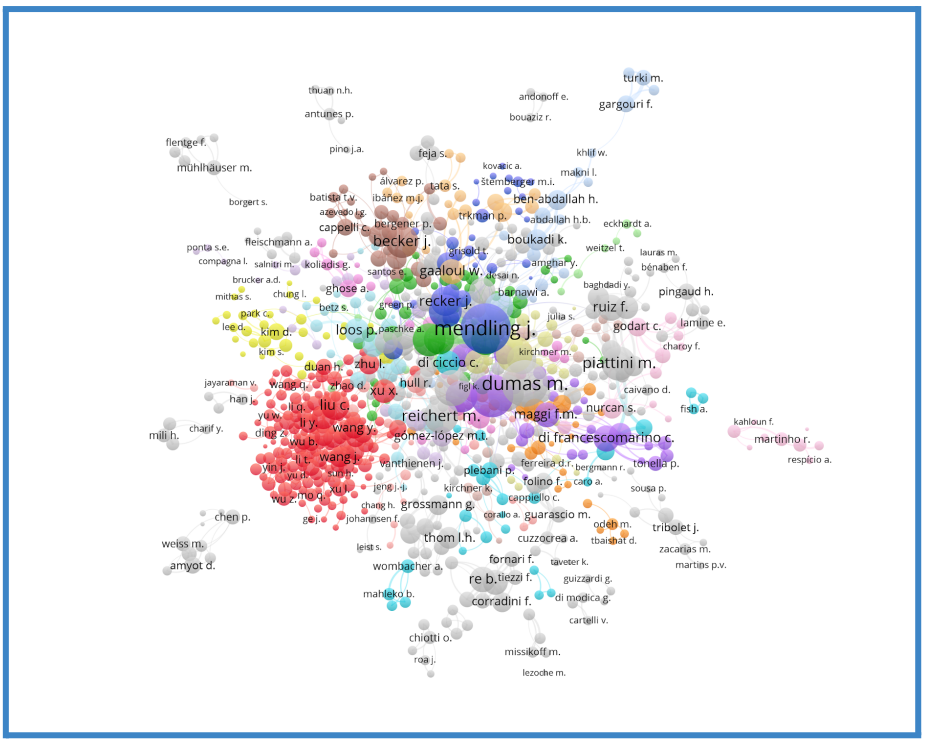


Figure 53. VOSviewer: Co-authorship (authors) network visualization (weight: TLS)

[Table 17] displays the top 10 most influential authors in the field, in descending order based on the TLS attribute. It is observed that the most influential authors consist of males only. Mendling has the highest amount of published documents (124) however does not have the highest amount of citations; Dumas ranks the highest in this regard (5168) and Mendling is following with 4911 citations. Mendling has also the highest amount of unique collaborations with 113 other authors and the highest TLS attribute. It is worth noticing that the 113 unique links of Mendling account for collaboration with 13% of the 840 authors who are contributing to the field. Since Mendling and Dumas rank the highest in terms of Documents, Citations and TLS, it would be expected that they would collaborate closely. In order to gain a first insight into their collaboration, the “min. Cluster size” in the action panel of the main window of VOSviewer was increased from 1 (resulting in 43 clusters) to 80 (resulting in 4 clusters), revealing that Mendling and Dumas do not belong to the same cluster [Figure 54].

Author	Links	TLS	Documents	Citations
mendling j.	113	315	124	4911
dumas m.	72	271	93	5168
la rosa m.	74	231	67	3261
weske m.	60	170	83	4609
weber b.	61	168	55	978
reijers h.a.	76	165	74	3589
piattini m.	17	148	60	1026
reichert m.	74	145	73	1549
van der aalst w.m.p.	55	139	57	4813
rosemann m.	66	135	59	2634

Table 17. VOSviewer: Co-authorship (authors)

Figure 54. VOSviewer: Co-authorship (authors) overlay visualization (weight: TLS)

In order to further investigate the interactions between the most influential authors, another network was created with different thresholds, which were selected in accordance with the [Table 17]: the "Minimum number of documents of an author" was set to 50 and the "Minimum number of citations of an author" was set to 1000 [Figure 55] while the weight attribute of the network was set again to "TLS".

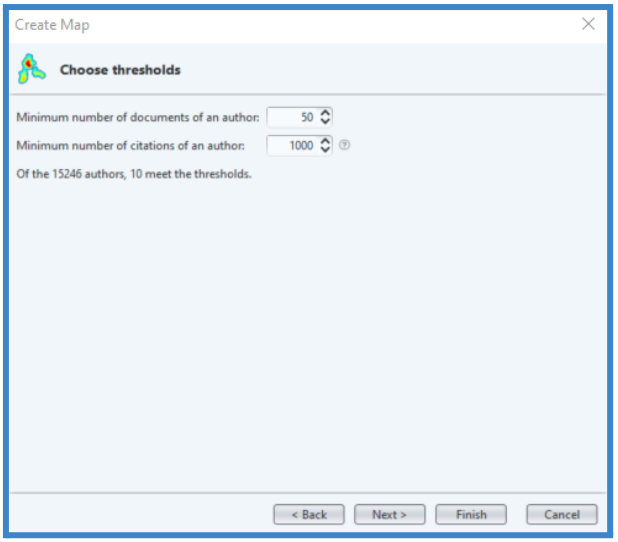


Figure 55. VOSviewer: Co-authorship (authors) window for selection of thresholds

In the produced network, it was directly obvious that the thickness of the line between Dumas and la Rosa indicated the highest Link strength with a value of 37, translating to the number of co-authored documents [Figure 56].

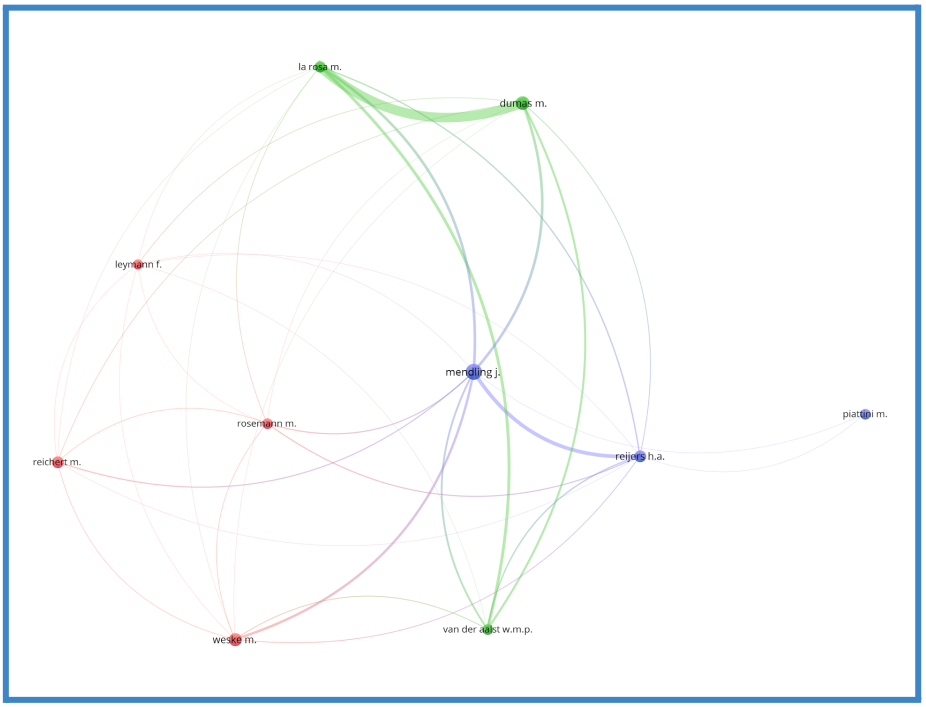


Figure 56. VOSviewer: Co-authorship (authors) network visualization of most influential authors (weight: TLS)

[Table 18] displays selected weight (i.e. Link strength, Links, TLS, documents) and score (i.e. average publication year) attributes of the map. The upper part of the table denotes the Link strength values (number of co-authored documents) between each pair of authors. Besides la Rosa and Dumas, the collaboration intensity of other authors is visible, with Mendling - Reijers having 14 co-authored documents, la Rosa - van der Aalst having 11 co-authored documents and Mendling - Dumas, as well as Mendling - la Rosa, having 10 co-authored documents.

Authors	la rosa	dumas	mendling	reijers	van der aalst.	weske	rosemann	reichert	leymann	piattini
la rosa										
dumas	37									
mendling	10	10								
reijers	5	3	14							
van der aalst	11	8	7	5						
weske	1	1	9	3	3					
rosemann	2	1	4	4	0	2				
reichert	1	2	4	1	0	2	2			
leymann	1	2	1	1	1	1	1	1		
piattini	0	0	1	1	0	0	0	0	0	
Links	8	8	9	9	6	8	7	7	8	2
TLS	68	64	60	37	35	22	16	13	9	2
Documents	67	93	124	74	57	83	59	73	51	60
Avg pub year	2015	2015	2013	2013	2012	2014	2012	2013	2009	2011

Table 18. VOSviewer: Co-authorship (authors) weight and score attributes

[Figure 57] exhibits the collaboration links for authors that have co-authored 7 or more documents (the “min strength” in the “options panel” was set to 7). The outcome of the analysis suggests that 6 out of the 10 selected authors collaborate closely and, more specifically, two collaboration patterns are observed: on the one hand, Mendling demonstrates distributed collaboration with the rest of the 6 authors, while, on the other hand, there is a closed network of collaboration between Dumas, la Rosa and van der Aalst.

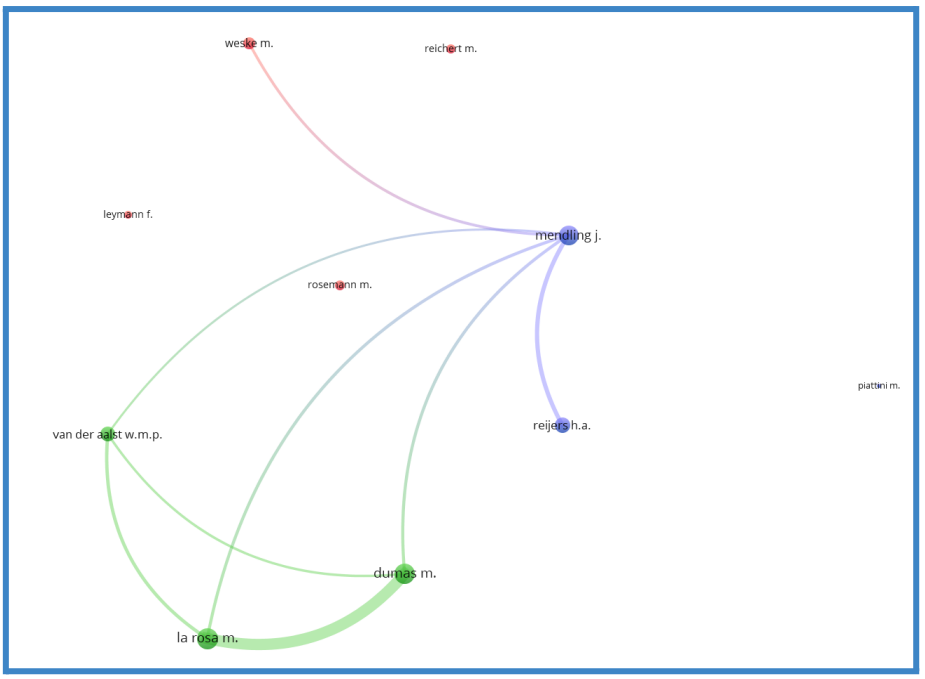


Figure 57. VOSviewer: Co-authorship (authors): of collaboration patterns

These two patterns were put under the microscope, with regard to their average publication year, with Dumas and la Rosa having the most recent average publication year (2015), while the publication year of the other pattern being 2013 [Figure 58].

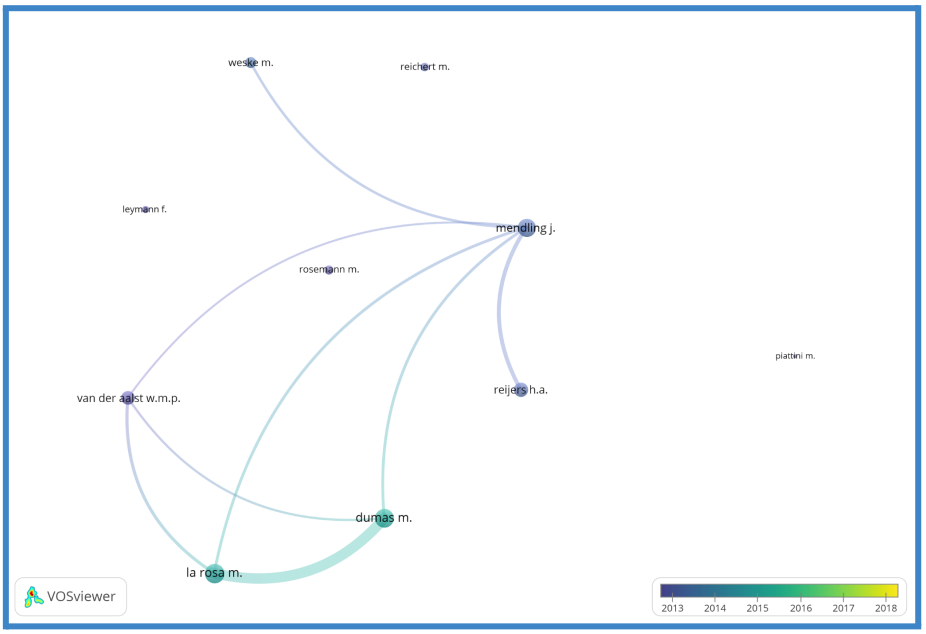


Figure 58. VOSviewer: Co-authorship (authors): average publication year of collaboration patterns

Besides the collaboration patterns, it was considered valuable to examine each author's dependency on the co-authorship by calculating the degree of collaboration with respect to the

total number of documents authored by each author. More specifically, the calculations were made based on the fraction

$$Dij = \frac{Link\ strenght\ of\ co-authorship}{author's\ documents}$$

[Table 19] showcases the results of the calculations for each author. The results highlight that 95% of la Rosa's publications are in collaboration with the other 5 authors, and 55% of his publications have been co-authored together with Dumas, in specific. Dumas and van der Aalst have 63% and 60% co-authorship rates with the other five authors, while their difference lies in the fact that Dumas has co-authored 40% of his publications together with la Rosa, while van der Aalst demonstrates a more diverse co-authorship tendency.

Author under examination	la rosa	dumas	mendling	reijers	van der aalst.	weske	total
la rosa		$\frac{37}{67} = 55\%$	$\frac{10}{67} = 15\%$	$\frac{5}{67} = 7\%$	$\frac{11}{67} = 16\%$	$\frac{1}{67} = 1\%$	95%
dumas	$\frac{37}{93} = 40\%$		$\frac{10}{93} = 11\%$	$\frac{3}{93} = 3\%$	$\frac{8}{93} = 9\%$	$\frac{1}{93} = 1\%$	63%
mendling	$\frac{10}{124} = 8\%$	$\frac{10}{124} = 8\%$		$\frac{14}{124} = 11\%$	$\frac{7}{124} = 5\%$	$\frac{9}{124} = 7\%$	40%
reijers	$\frac{5}{74} = 7\%$	$\frac{3}{74} = 4\%$	$\frac{14}{74} = 19\%$		$\frac{5}{74} = 7\%$	$\frac{3}{74} = 4\%$	41%
van der aalst	$\frac{11}{57} = 19\%$	$\frac{8}{57} = 14\%$	$\frac{7}{57} = 12\%$	$\frac{5}{57} = 9\%$		$\frac{3}{57} = 5\%$	60%
weske	$\frac{1}{83} = 1\%$	$\frac{1}{83} = 1\%$	$\frac{9}{83} = 11\%$	$\frac{3}{83} = 4\%$	$\frac{3}{83} = 4\%$		20%

Table 19. VOSviewer: Co-authorship (authors): authors' dependencies on the co-authorship

The results of the co-authorship analysis indicate that out of the 841 authors that engage in collaborations, only six of them collaborate more intensively, and all of them are male. These six authors rank the highest among all in terms of produced publications and received citations and two collaboration patterns are formed among them, with 50 and 56 co-authored documents published by each group of authors respectively. These numbers account for a significant amount of contributions to the research field. In this context, it was observed that some of these authors rely on their collaborations to a great extent for producing publications. Furthermore, the average publication years (between 2013-2015) which are considered fairly recent, make these contributions highly relevant to the direction that the field is taking today. These observations lead to the conclusion that the research field is highly influenced by these authors, who set the direction that the field is taking until today.

5.2. Co-authorship analysis (Countries)

The next type of analysis was co-authorship analysis, with “Countries”, as the unit of analysis. As mentioned in Chapter 2.1 “When authors from two or more countries contribute to a given article, the authors’ countries are considered as collaborating countries.”

For this type of analysis, the unified CSV file was used as input. The option to “ignore documents co-authored by a large number of countries” was disabled [Figure 59] and the fields “minimum number of documents of a country” and “minimum number of citations of a country”, were left with their default value; 5 and 0 respectively [Figure 60]. As indicated by the software, out of the 291 countries, only 84 met these thresholds and out of these 82 were connected (the remaining 2 did not have any links) [Figure 61].

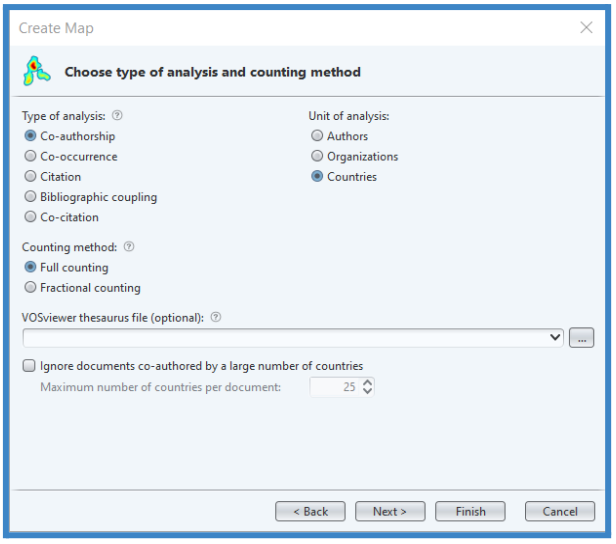


Figure 59. VOSviewer: window for selection of Co-authorship as the type of analysis with Countries as the unit of analysis

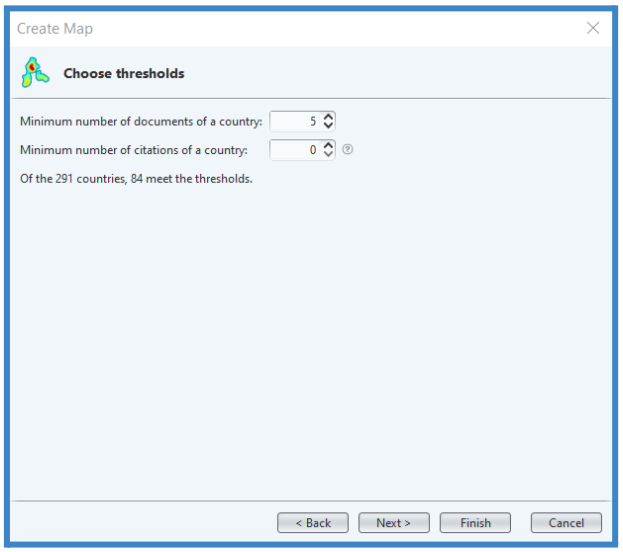


Figure 60. VOSviewer: Co-authorship (countries) window for selection of thresholds

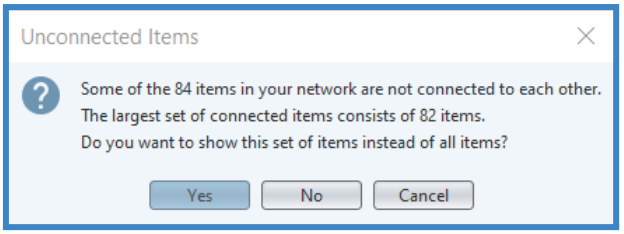


Figure 61. VOSviewer: Co-authorship (countries) warning message for connected items

The network was created based on the weight attribute "TLS" [Figure 62]. The nodes represented the countries and their size was proportional to the TLS of each country (bigger size denoted higher TLS). The links between countries indicated the existence of collaboration (co-authorship). Each link between countries had a "**Link strength"** attribute, denoting the times that they have collaborated on publications (i.e. the total number of co-authored documents). When looking at each country individually, the unique number of countries that it has collaborated with was reflected in its "**Links"** attribute (a maximum number of links could be 81, suggesting that the country has collaborated with all the countries present on the map). The "**TLS"** attribute of a country, which was reflected on the node size, indicated the cumulative "**Link strength"** of all of its "**Links"**. Additionally, each country had "**documents**" and "**citations**" as weight attributes.

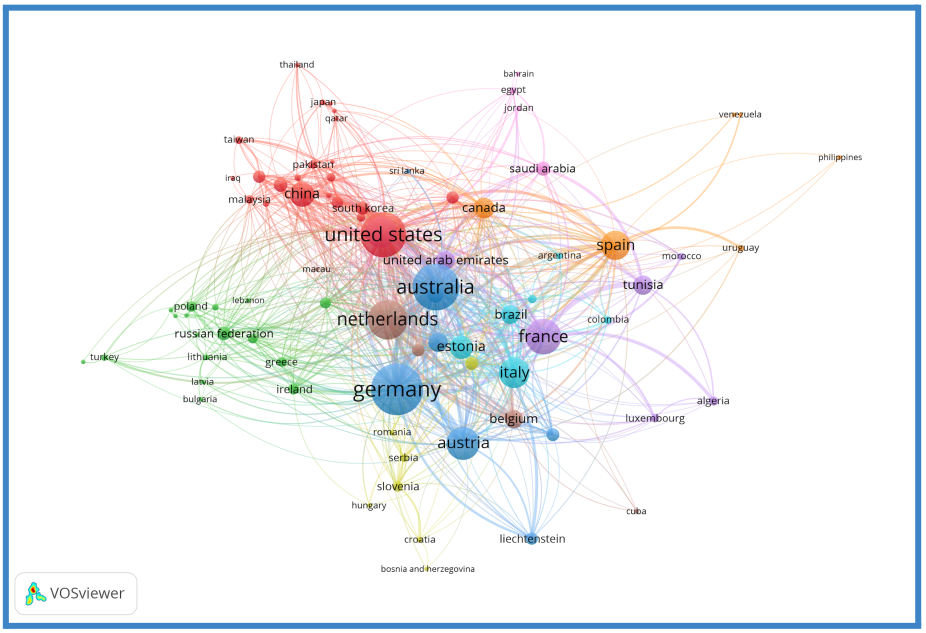


Figure 62. VOSviewer: Co-authorship (countries) network visualization (weight: TLS)

[Table 20] exhibits the ten most influential countries according to each weight attribute in descending order. Initially, it was observed that these countries belong to the ones with the highest GDP worldwide (International Monetary Fund, 2022) which implies a high investment in the research of business processes. More specifically, Germany and the US are the countries

with the highest weight attributes among others. They have both the highest production of **documents** (over 1.000) and, foreseeably, the highest amount of **citations** (over 20.000). In terms of their unique collaborations (**Links**), the US has 58 unique collaborations out of the 82 countries in the network (71%) while Germany has 67% diversity rate of collaboration with other countries. Lastly, Germany has the highest intensity of total collaborations (**TLS**=758) with the US being third in the ranking.

Another deviation between Germany and the US is observed in the overlay visualization, which indicated that the average publication year of documents originated in the US is 2010 (blue shade) while for Germany is 2012 (green shade) [Figure 63]. It would be expected that these two countries collaborate closely, however, the cluster density visualization implied that the countries belong in different clusters: Germany is clustered together with Austria and Australia while the US is clustered with China [Figure 64].

country	documents
germany	1788
united states	1091
china	862
australia	799
italy	618
united kingdom	526
netherlands	495
france	473
austria	449
spain	359

country	citations
germany	26038
united states	22829
australia	18468
netherlands	15890
united kingdom	9120
austria	7823
italy	7446
estonia	5378
china	5128
france	4305

country	Links
united states	58
united kingdom	57
germany	55
australia	52
france	46
netherlands	44
canada	43
spain	42
italy	38
austria	37

country	TLS
germany	758
australia	629
united states	614
netherlands	486
france	402
austria	360
united kingdom	324
italy	312
spain	292
china	221

Table 20. VOSviewer: Co-authorship (countries) weight attributes of most influential countries

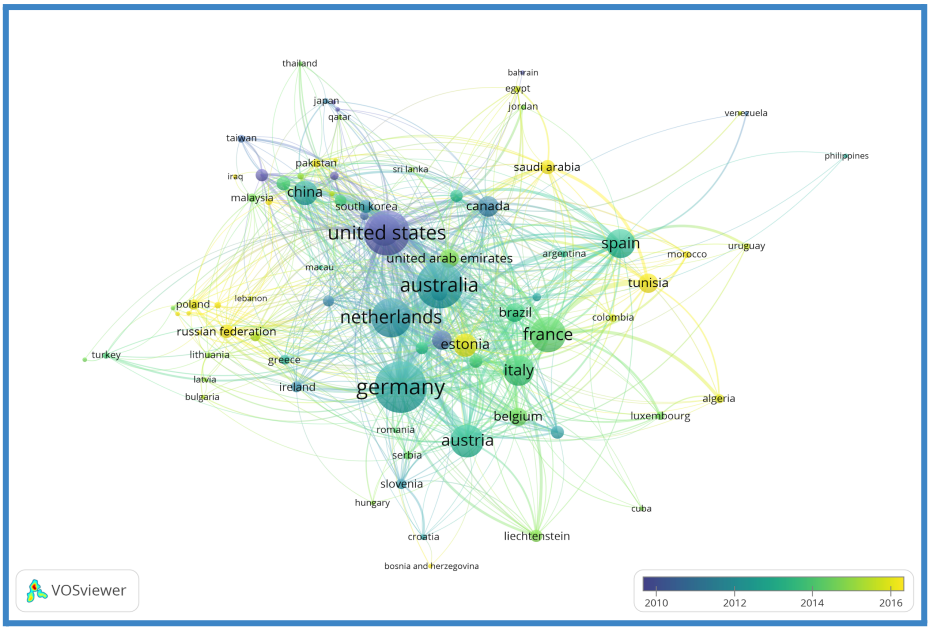


Figure 63. VOSviewer: Co-authorship (countries) overlay visualization (weight: TLS)

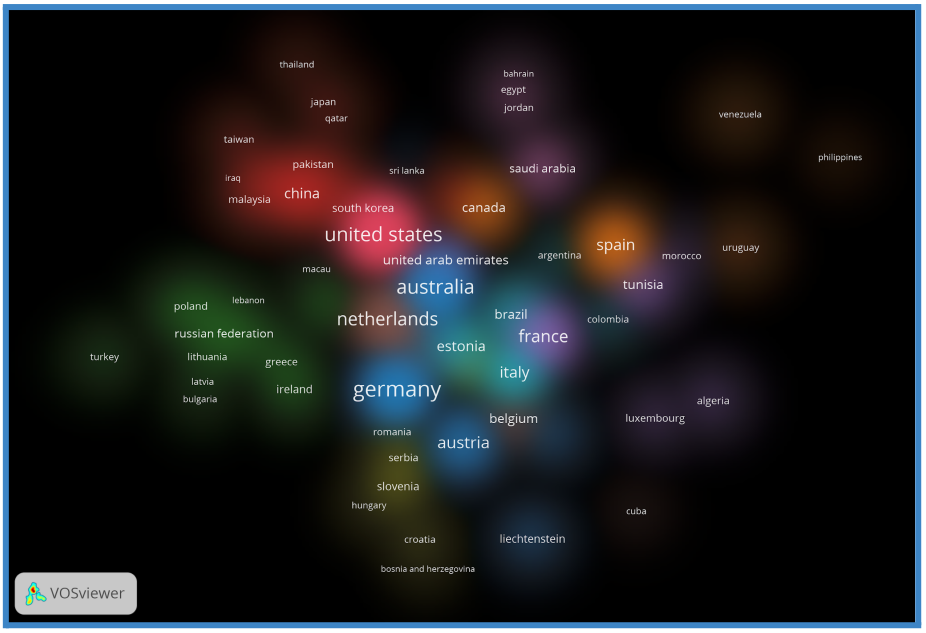


Figure 64. VOSviewer: Co-authorship (countries) density visualization (weight: TLS)

The last observation was further explored with the creation of a network consisting of the top 10 countries in terms of documents and citations (300 and 4000, respectively) [Figure 65]. In the produced network all of the countries co-author documents with each other (except for China and Spain which do not share a link) [Figure 66]. The TLS of each country was further explored (i.e. the total volume of co-authored publications with the rest of the 9 countries) and was compared with each country's TLS from the previous step (i.e. cumulative co-authored publications with the rest of the 81 countries). As a comparison metric, the ratio between the two

TLS attributes was calculated and the results indicated that, on average, more than half of the produced documents of each country were co-authored with one of the 9 countries [Table 21].

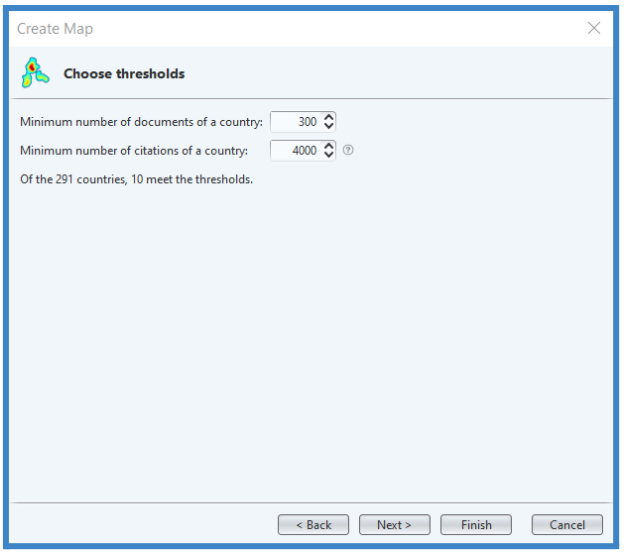


Figure 65. Figure 52. VOSviewer: Co-authorship (countries) window for selection of thresholds

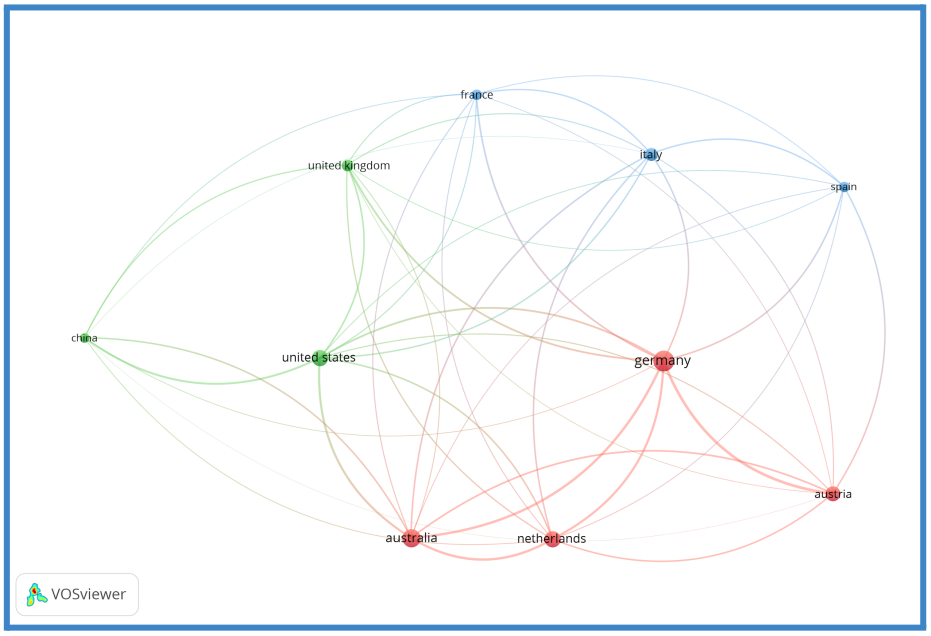


Figure 66. VOSviewer: Co-authorship (countries): Links between 10 most influential countries

country	TLS with 10 countries	TLS with 82 countries	ratio
germany	458	758	60%
australia	370	629	59%
united states	301	614	49%
netherlands	301	486	62%
austria	255	360	71%
italy	209	312	67%
united kingdom	176	324	54%
france	146	402	36%
spain	141	292	48%
china	137	221	62%

Table 21. VOSviewer: Co-authorship (countries)

In order to spot the countries with the highest collaboration rate, the "Min. strength" was adjusted in the "Options panel", which had the default value set to 0. Firstly, the value was set to 100, which showcased that the highest link strength existed between Germany and Austria [Figure 67]; as opposed to the expectation of seeing the strongest link between Germany and the US.

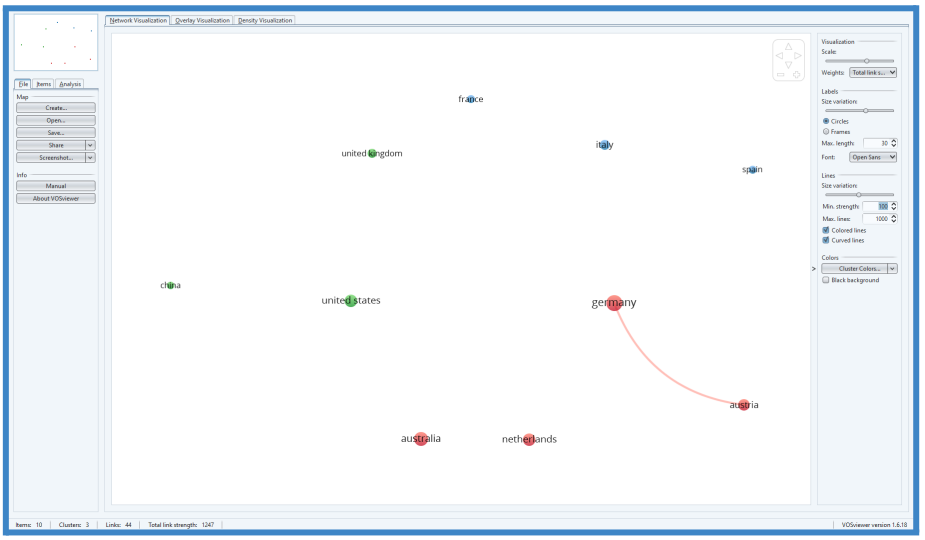


Figure 67. VOSviewer: Co-authorship (countries): Links for Min. strength=100

The "Min. strength" was subsequently set to 80 which revealed strong links between Germany and Australia but also Australia and the Netherlands [Figure 68]. Lastly, when the value was set to 60 it exhibited links between the US and Australia and between Australia and the Netherlands [Figure 69]. [Table 22] summarizes all the link strength attributes among these 10 countries.

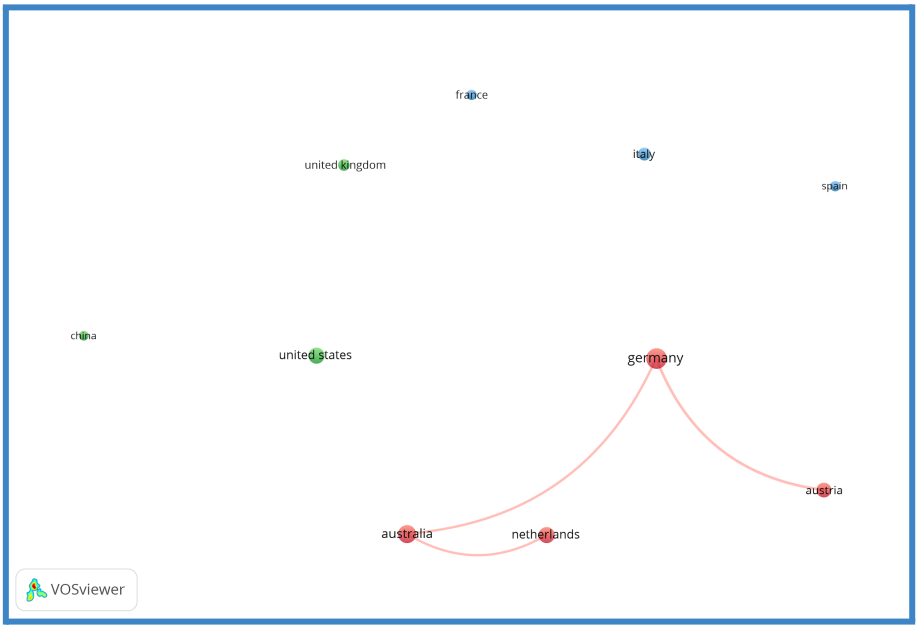


Figure 68. VOSviewer: Co-authorship (countries): Links for Min. strength=80

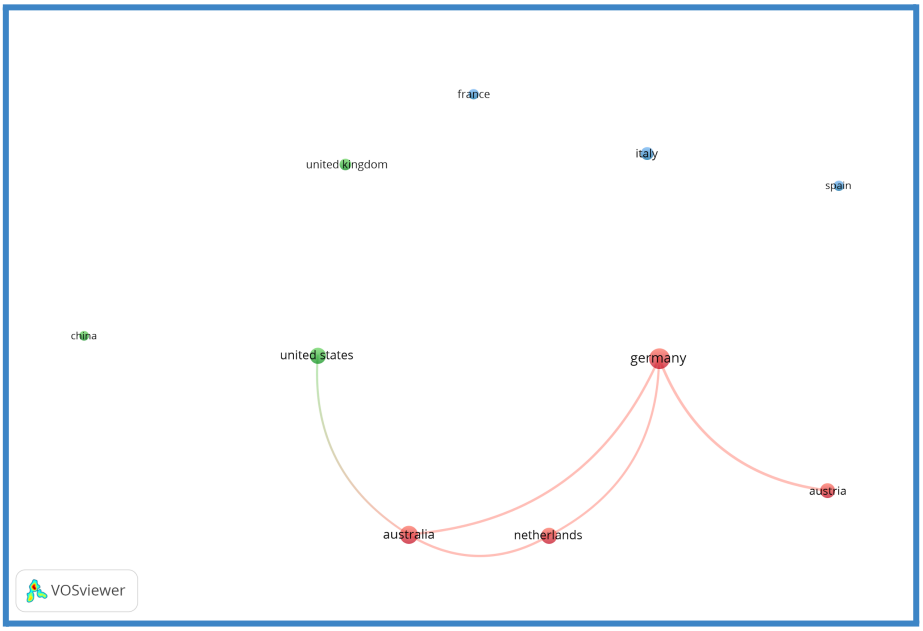


Figure 69. VOSviewer: Co-authorship (countries): Links for Min. strength=60

country	DE	AU	US	NL	AT	IT	UK	FR	ES	CN
DE										
AU	84									
US	47	62								
NL	70	80	33							
AT	103	43	17	34						
IT	31	33	27	31	13					
UK	45	11	35	22	6	15				
FR	38	15	15	12	8	26	13			
ES	33	9	10	10	30	31	8	10		
CN	7	33	55	9	1	2	21	9	0	

Table 22. VOSviewer: Co-authorship (countries) Link strength between 10 most influential countries

From the Co-authorship analysis with countries as the unit of analysis occurs that countries with high GDP invest a lot into the research of business processes. More specifically, it was found that the US and Germany are the most influential countries in the research field, however, do not share the highest volume of co-authored publications. It is worth mentioning that the GDP of these countries [Table 23] indicates that during the past years they mainly operate in the services industry (with the proportion of their GDP being over 60%) and not in the industry or manufacturing. Thus it is understood that the investment in research in business processes puts the services into the spotlight.

In the effort of deconstructing the collaboration patterns, the collaboration of Germany with Austria and the Netherlands was considered reasonable due to them being neighbouring countries which operate within the European Union and share similar distribution and growth in terms of the GDP of each sector (The World Bank, 2022). The insights from Scopus analytics provide an additional perspective on the topic, indicating that the top funding sponsors were European; with the top funding sponsor being the European Commission [Figure 70].

The collaboration between Australia with Germany could be explained based on their differences in the development of their sectors; Australia has higher development in Agriculture services compared to Germany, while Germany demonstrates higher development in the industry and manufacturing sectors. Respectively, the collaboration between the US and Australia could be based on Australia's higher development percentages in the sectors of agriculture and industry and the State's development in the manufacturing and services sector. Thus these collaborations are assumed to be formed on the basis of knowledge sharing in different sectors.

All in all, the observations led to the conclusion that two countries lead the research output of the field; which on the one hand can be beneficial for underdeveloped countries, however on the other hand the research field is confined to the knowledge and perspectives of these countries.

	GDP		Agriculture		Industry		Manufacturing		Services	
	$ billions		% of GDP		% of GDP		% of GDP		% of GDP	
Country / Year	2010	2020	2010	2020	2010	2020	2010	2020	2010	2020
Germany	3,396.4	3,846.4	0.8	0.7	26.8	26.5	19.7	18.2	62.3	63.3
United States	14,992.1	20,953.0	1.0	0.9	19.4	18.2	11.9	10.9	76.2	77.3
Australia	1,147.6	1,327.8	2.2	2.0	25.2	25.5	8.0	5.7	65.7	66.3
Austria	391.9	433.3	1.3	1.1	25.5	25.5	16.5	16.3	62.3	63.1
Netherlands	846.6	913.9	1.8	1.6	19.7	17.8	10.5	10.8	68.4	69.8

Table 23. World Development Indicators: Structure of output (The World Bank, 2022)

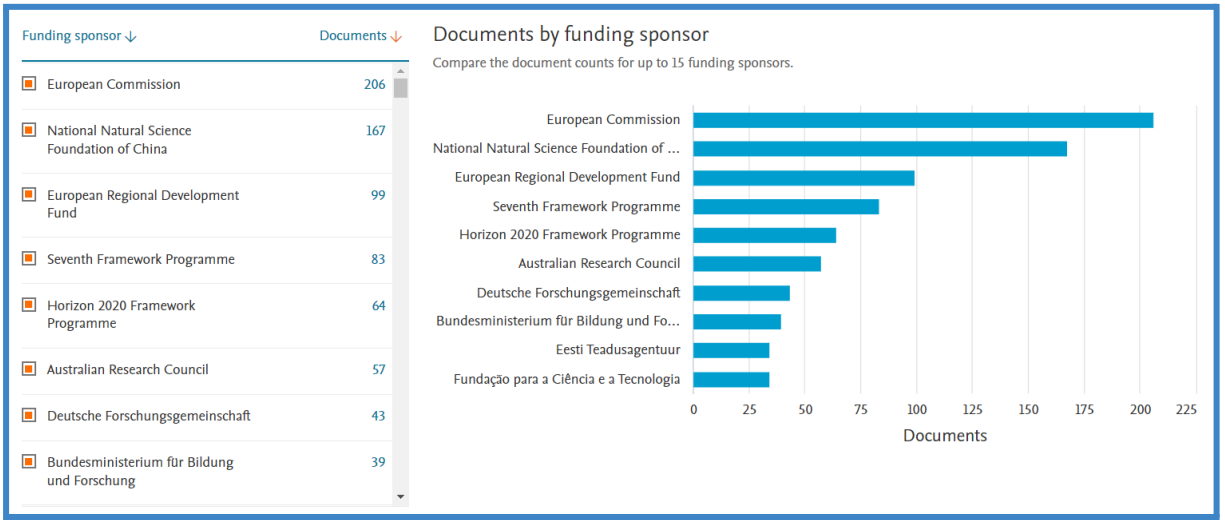


Figure 70. Scopus analytics: funding sponsors

5.3. Co-occurrence analysis (Author keywords)

The last type of analysis performed was the Co-occurrence analysis which lies on the assumption that "words that frequently appear together have a thematic relationship with one another." For the Co-occurrence type of analysis, the "Author keywords" were selected as the unit of analysis [Figure 71] since they were considered the most concise, still reliable source which reflected the content of the publications. The unified CSV file was used as input and the "Minimum number of

occurrences of a keyword" was left with its default value of 5, resulting in 986 keywords which comprised the final network [Figure 72].

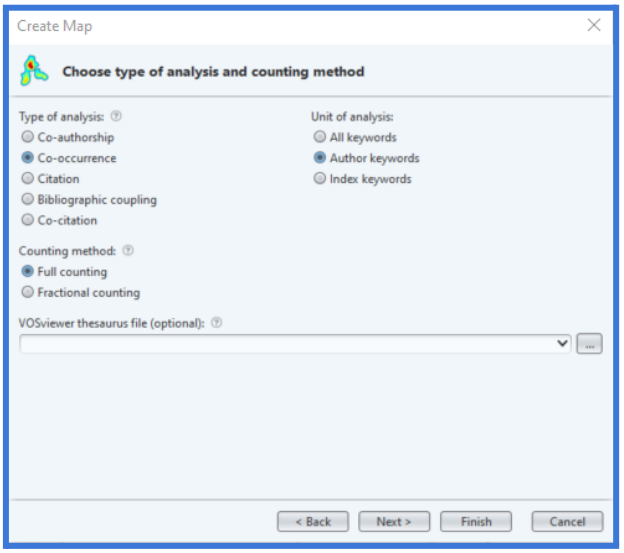


Figure 71. VOSviewer: window for selection of Co-occurrence as the type of analysis with Author keywords as the unit of analysis

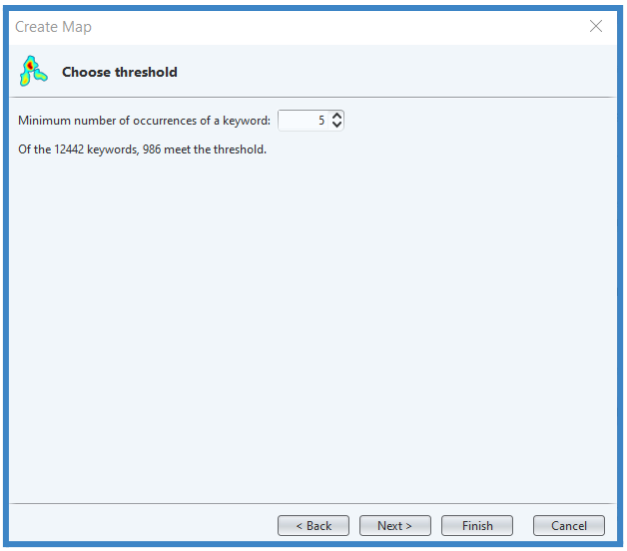


Figure 72. VOSviewer: Co-occurrence window for selection of threshold

The network was created based on the weight attribute "Occurrences". The nodes represented the Author keywords thus their size was proportional to the TLS of each term (bigger size denoted higher TLS). For a given keyword, the Links attribute indicated the unique number of co-occurrences of a term with each of the rest of the terms (the highest number of links would be 985, meaning that the keyword appeared every time another word from the network was used as a keyword). Between 2 terms, the link strength indicated the number of publications in which the two terms appeared together (co-occurred). Ultimately, the TLS attribute of each term suggested the cumulative link strength of all the links that a term had with other terms. The

produced network is displayed in [Figure 73] and the overlapping central circles are shown in detail in [Figure 74].

The keywords with over 1.000 occurrences were “business process”, “business process management” and “business process modeling” [Table 24]. As aforementioned, the highest number of links a keyword could have was 985, thus it is clear that the “business process” is a co-occurring keyword 77% of the times that another keyword appears. Similarly, “business process management” has 74% and “business process modeling” 65% co-occurring possibility. Of these keywords, the one which appears in the bibliography more recently (2017) is “process mining” while the “business process reengineering” keyword has the oldest average publication year (2011).

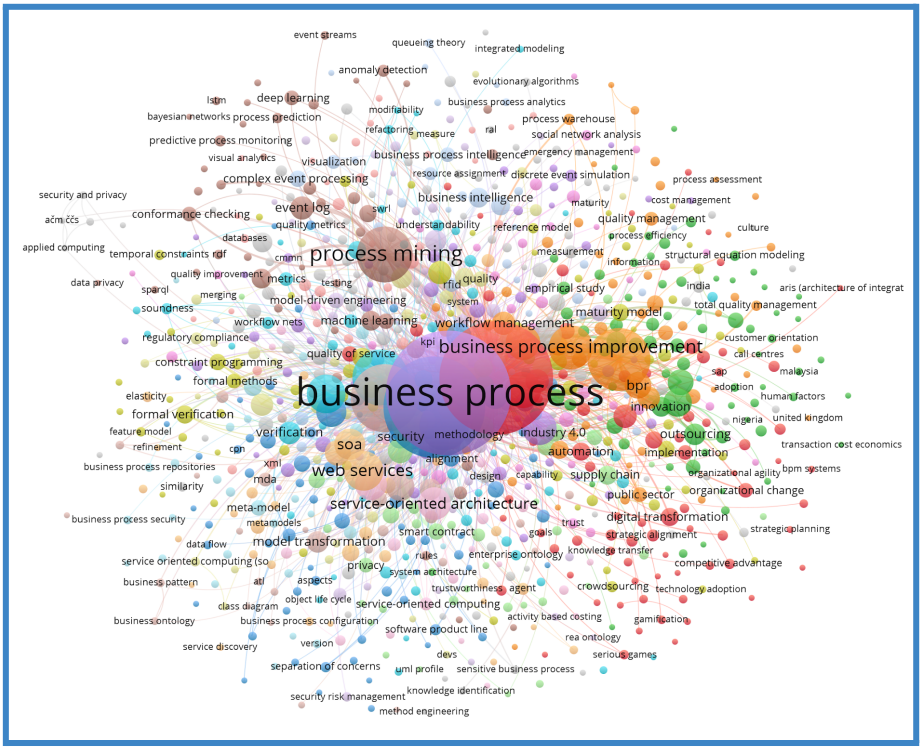


Figure 73. VOSviewer: Co-occurrence network of Author Keywords (weight: Occurrences)

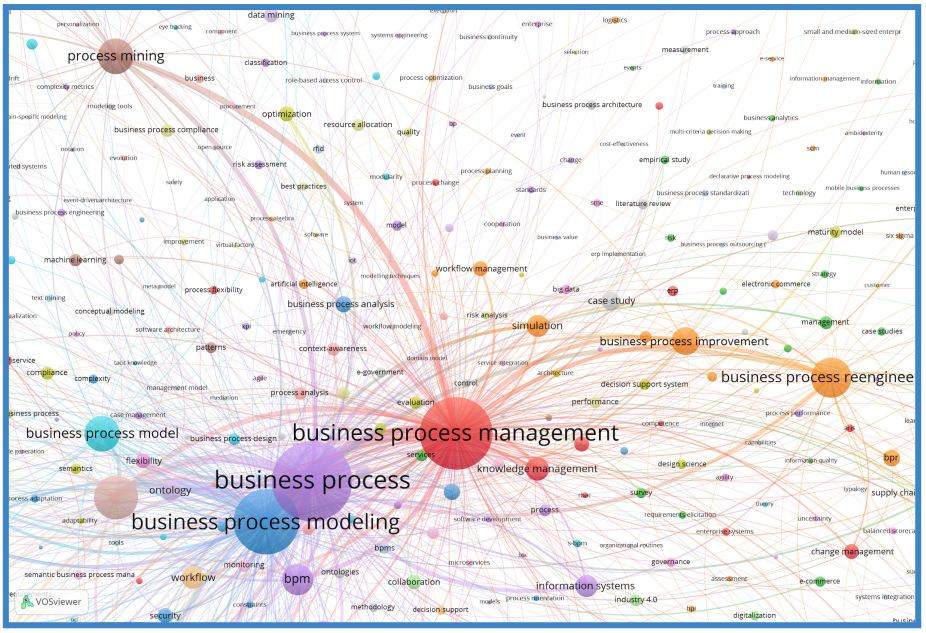


Figure 74. VOSviewer: Co-occurrence: Zoom in the keywords with the highest TLS

+

Keyword	Links	TLS	Occurrences	Avg pub year
business process	757	3525	1685	2014
business process management	728	3118	1391	2014
business process modeling	639	2458	1117	2013
business process modeling notation	405	1321	505	2015
business process reengineering	294	845	432	2011
process mining	250	753	338	2017
business process model	311	710	341	2015
bpm	294	677	259	2014

Table 24. Co-occurrence weight attributes

The co-occurrence ratio between the mostly occurred keywords was further explored with the increase of the “Min. strength” option to 100 [Figure 75].

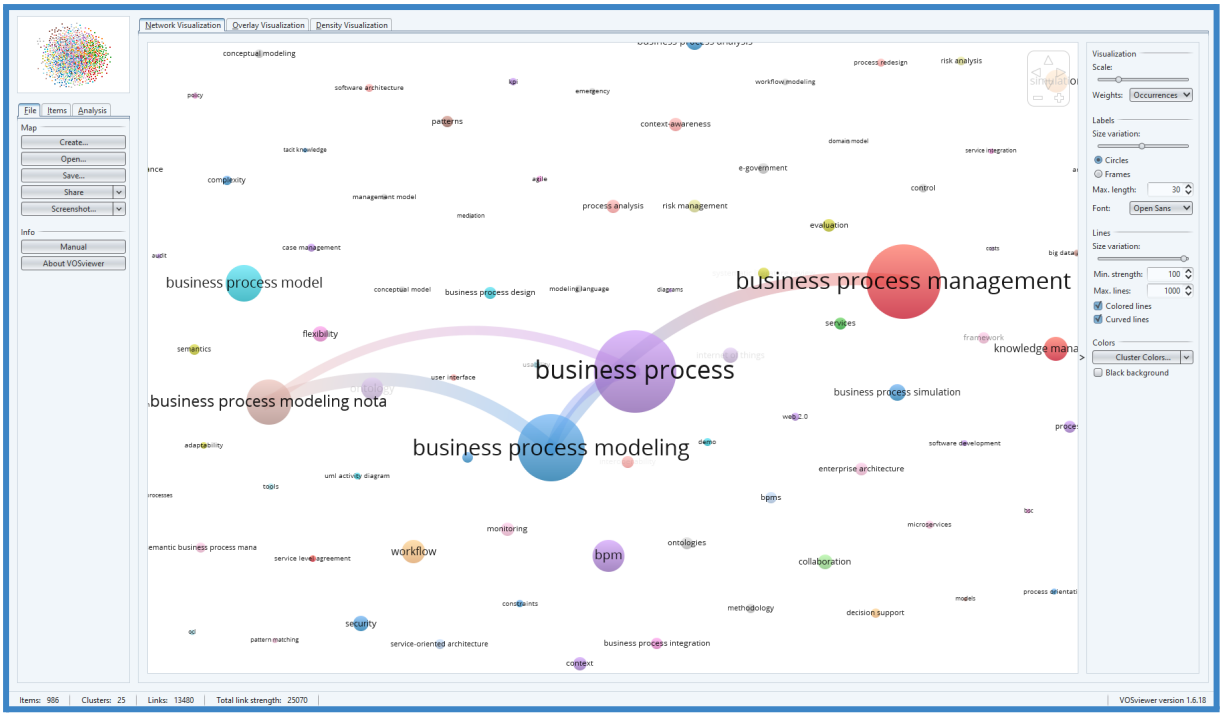


Figure 75. VOSviewer: Co-occurrence network with Min. strength = 100

The Link strength between the resulting, connected keywords reflects the number of documents (out of the 10.506 documents) in which they appear together and is presented in [Table 25]. The results indicate that the terms which co-occur the most are “business process modeling” with “business process modeling notation” and “business process management” with “business process modeling”. Interestingly, the terms “business process” and “business process management” did not seem to co-occur as often as expected thus the former decision to replace the term “business process management” keyword from the search query and replace it with “business process” was proved to be correct. The results indicate that “business process” is more likely to be connected with “business process modeling” than with “business process management”.

Term 1	Term 2	Link strength
business process modeling	business process modeling notation	148
business process management	business process modeling	141
business process	business process modeling	110
business process	business process modeling notation	101
business process	business process management	72

Table 25. VOSviewer: Co-occurrence link strength between keywords

In order to further explore the trends in the field of business processes, the words with an average publication year between 2019 - 2021 were spotted. The ones with over 50 occurrences

are presented in [Table 26]. The trends in the field seem to be In alignment with the trends in the sectors of industry (industry 4.0, digital transformation) and technology (blockchain, machine learning, artificial intelligence). For each of these keywords, the highest link strengths with the respective terms are displayed. It is noted that all of the keywords co-occur with “business process” and “business process management”.

Ultimately, the Co-occurrence analysis indicated that the most representative keywords of the field of business processes are extensions of the term “business process” (i.e. business process management, modeling, reengineering, etc). Based on the co-occurrence link strengths it was visible that similar words are not interchangeable; “business process” does not frequently co-occur with “business process management” and respectively “business process modeling” does not frequently co-occur with “business process modeling notation”. That implies that terminology in the field is specific and should be used accordingly.

Furthermore, it was noticed that 2014-2015 were the years that “business process” was widespread as the central research topic. The current research trends seem to follow the global trends in industry and technology, however, the occurrences of the respective keywords are significantly lower than expected.

In conclusion, the main topics of the research field revolve around the contemporary business process approaches and newly emerging trends are still not incorporated significantly into the research of the field.

Term 1	Avg pub year	Occurrences	Links	Term 2	Link strength
blockchain	2020	70	91	business process management	26
				business process	17
				smart contract	26
				ethereum	5
machine learning	2019	43	68	business process management	10
				business process	8
				artificial intelligence	6
				process mining	8
digital transformation	2020	38	55	business process management	12
				business process	9
industry 4.0	2021	36	67	business process management	13
				business process	6
				digitalization	6
smart contract	2020	30	40	business process management	11
				blockchain	26
digitalization	2020	27	57	business process management	5
				business process	8
				Industry 4.0	6
				digital transformation	6
artificial intelligence	2019	27	50	business process management	4
				business process	6
				machine learning	6

Table 26. VOSviewer: Co-occurrence of keywords with recent average publication year

6. Conclusion

6.1. Main remarks

As mentioned in the first chapter of the thesis its primary aim was to examine the publication output regarding “business processes” until 2022.

Chapter 2 presented the comparison of the main review methods which concluded with Bibliometric Analysis as the one which properly aligned with the broad scope and the amount of literature required in order to fulfil the aim of the study. Bibliometric Analysis was deconstructed and a flowchart with comprehensive implementation steps was selected from the bibliography as the basis of the analysis. Upon experimentation, a complementary preprocessing procedure was also proposed as complementary to the main procedure. The suggested approach aimed to limit the inevitable reiterations of the main Bibliometric Analysis process with the acquisition of insights that would ensure a better understanding of the possibilities as well as the limitations of the Bibliometric Analysis.

In Chapter 3 the proposed Bibliometric Analysis preprocessing was deployed and a step-by-step analysis and application of each step was performed. The four steps consisted of the examination of Bibliometric Analysis techniques, the examination of the field under study, the examination of various databases and the examination of available software. At the end of this chapter, a holistic view of the bibliometric techniques and tools was gained.

Chapter 4, addressed the Bibliometric Analysis process which was acquired from the bibliography. Similarly to the previous chapter every step of the procedure was initially analyzed thoroughly and then deployed. The four steps of this process consisted of the aim and scope definition, the technique selection, the data collection and the data analysis. The chapter ended with all the necessary preparation of the dataset, in order to be further exploited.

Chapter 5 exhibited the results of the Co-authorship (authors), Co-authorship (countries) and Co-occurrence analysis with the utilization of the software VOSviewer. The results were able to satisfy the objectives and the research questions of the thesis.

6.2. Research contribution

The primary contribution of the present thesis accounts for the thorough analysis of the whole Bibliometric Analysis process. The extant literature of the field under study mainly focuses on

results with a brief reference on the whole Bibliometric Analysis procedure details on the utilized techniques. The thesis can serve as a simple, yet thorough guide for the profound understanding of the Bibliometric Analysis principles and application.

More specifically, the tables of Science Mapping techniques summarized the collective knowledge acquired from various sources. As a matter of fact, the deployment of such techniques in the business field is little, compared to the Performance Analysis techniques and the relevant descriptions of these methods seemed complex and not very detailed. The tables aim to provide a comprehensive description of these techniques which can consequently encourage researchers to deploy them more frequently.

In addition to the main procedure, a complementary Bibliometric Analysis preprocessing procedure is proposed. The procedure was a combination of research on the Bibliometric Analysis with empirical findings. Its aim is to help novice researchers and researchers with no prior knowledge of bibliometrics to gain insights into the necessary methods and tools that should be taken into account before proceeding with the main procedure.

In terms of contributions to the field itself, the thesis aimed to examine the degree to which the field is shaped by diverse contributions and multidimensional aspects. It was found that the field is more bounded than diverse with regard to the contributing authors and their countries. Strong collaborations are observed among highly influential authors whose publications influence the direction of the field until today. At the same time, developed countries invest highly in business process research which can be considered a limiting factor for the acquisition of insights from the operation of businesses in other countries. All in all the research of the field can be considered rich however not multidimensional and inclusive.

The last objective of the thesis regarded the discovery of the trends in the field and it was noticed that the research mainly revolved around business process-related topics that bloomed in 2015. There is research addressing the combination of these topics with advances in industry and technology however the volume of such publications is still very low.

6.3. Research limitations & future work

The limitations in the present thesis accounted for the selection of the final dataset. Due to the formulation of the query Scopus was chosen as the database, however other databases seemed equally promising. It would be interesting to compare the results from different databases.

At the same time, VOSviewer posed the limitation of compatibility with certain databases and additional confinement on the use of input files from one database only. Another future approach that could be adopted could be to collect datasets from different databases, join them into a single file and use another software that would support the analysis of the results. With this approach, a more representative amount of literature would be used and would provide a more holistic, yet profound understanding of the field.

Lastly, the Citation, Co-citation and Co-occurrence techniques could be exploited in order to gain additional insights into the dynamics of the field.

6.4. Conclusion

The motivation of the present thesis originated in the changes that are observed in the business arena during the times of the pandemic. These changes are not reflected in the research of business processes until 2022 however it is yet to be seen whether the field of business processes is saturated or the implications of the global phenomena are still under investigation and new approaches to business processes will emerge over time.

REFERENCES

Ali, J., Jusoh, A., Idris, N., Qureshi, N. A., Nasrullah, Abbas, A. F., & Nabeel Nisar. (2022). Contributions, links, association strength and collaborations in SERVQUAL & healthcare research. *Journal of Public Value and Administrative Insight*, *4*(4), 350–363. Readers Insight. https://doi.org/https://doi.org/10.31580/jpvai.v4i4.2408

Dimensions Release Notes. (2020, April 20). Dimensions. https://www.dimensions.ai/release-notes/#:~:text=Dimensions%20users%20can%20download%20result

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*(0148-2963), 285–296. Elsevier. https://doi.org/10.1016/j.jbusres.2021.04.070

Donthu, N., Kumar, S., & Pattnaik, D. (2020). Forty-five years of Journal of Business Research: A bibliometric analysis. *Journal of Business Research*, *109*, 1–14. Elsevier. https://doi.org/10.1016/j.jbusres.2019.10.039

Fahimnia, B., Sarkis, J., & Davarzani, H. (2015). Green supply chain management: A review and bibliometric analysis. *International Journal of Production Economics*, *162*(0925-5273), 101–114. Elsevier. https://doi.org/10.1016/j.ijpe.2015.01.003

International Monetary Fund. (2022). *GDP, current prices*. IMF. https://www.imf.org/external/datamapper/NGDPD@WEO/OEMDC/ADV

Lens Export Results. (2021, July 30). Lens Support. https://support.lens.org/knowledge-base/export-results/

Lizano-Mora, H., Palos-Sanchez, P. R., & Aguayo-Camacho, M. (2021). The Evolution of Business Process Management: A Bibliometric Analysis. *IEEE Access*, *9*, 51088–51105. https://doi.org/10.1109/access.2021.3066340

Manzari, L. (2021, November 18). *Library Guides: Bibliometrics: Tools and Software*. LIU Post; Long Island University. https://liu.cwp.libguides.com/c.php?g=225325&p=4966525

Nousias, N. (2021). *Business Process and Decision automation: End-to-end deployment with a BPMN and DMN-based workflow engine* [MSc Thesis].

OpenRefine. (2012). OpenRefine. https://openrefine.org

OpenRefine user manual. (2022). OpenRefine. https://docs.openrefine.org/

Ortigueira-Sánchez, L. C., & Risco-Martínez, S. L. (2021). The situation of management research in Peru. A bibliometric analysis in the Bicentennial. *Journal of Business*, *13*(2), 72–87. https://doi.org/10.21678/jb.2021.1729

Persson, O. (n.d.). *BibExcel*. University of Vienna. Retrieved April 5, 2022, from https://homepage.univie.ac.at/juan.gorraiz/bibexcel/

Persson, O., Danell, R., & Wilborg Schneider, J. (2009). How to use Bibexcel for various types of bibliometric analysis. *Celebrating Scholarly Communication Studies: A Festschrift for Olle Persson at His 60th Birthday*, 9–24.

PubMed User Guide. (2022, April 6). PubMed. https://pubmed.ncbi.nlm.nih.gov/help/

Sajovic, I., & Boh Podgornik, B. (2022). Bibliometric Analysis of Visualizations in Computer Graphics: A Study. *SAGE Open*, *12*(1). https://doi.org/10.1177/21582440211071105

The World Bank. (2022, May 25). *World Development Indicators: Structure of output*. Worldbank. http://wdi.worldbank.org/table/4.2

Upgrades to Scopus: Export Limits Increased to 20,000. (2013, July 20). Scopus Blog. https://blog.scopus.com/posts/upgrades-to-scopus-export-limits-increased-to-20000

Van Eck, N. J., & Waltman, L. (2022). VOSviewer Manual. In *VOSviewer*. https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.18.pdf

Vergidis, K. (2008). *Business process optimisation using an evolutionary multi-objective framework* [Ph.D. dissertation].

VOSviewer. (2022). VOSviewer; Leiden University. https://www.vosviewer.com/

Web of Science: Limit on exporting. (2022, April 20). Clarivate. https://support.clarivate.com/ScientificandAcademicResearch/s/article/Web-of-Science-Limit-on-exporting?language=en_US