Chapter 10 / Capítulo 10

Applied bibliometrics. From data to publication (English Edition)
ISBN: 978-9915-9680-7-0
DOI: 10.62486/978-9915-9680-7-0.ch10

Academic Writing / Redacción Académica

The culmination of any bibliometric research lies in its ability to communicate findings clearly, rigorously, and persuasively. This chapter focuses on the practical aspects of academic writing, structuring articles, and effectively communicating bibliometric results to different audiences. From the initial conceptualization of the study to the final presentation of results, each chapter provides concrete tools for transforming data analysis into meaningful scientific contributions that enrich our understanding of the dynamics of knowledge.

10.1. The idea of bibliometrics

A bibliometric study differs from other forms of research in its quantitative approach to analyzing the production, dissemination, and impact of scientific knowledge through specific metrics and statistical techniques. The conception of bibliometric research, like all scientific research, begins with the identification of a gap in the understanding of the structure or dynamics of a scientific field, where the analysis of patterns in the literature can offer solutions. Unlike systematic reviews that synthesize substantive findings or qualitative studies that explore meanings and experiences, bibliometrics focuses on macroscopic patterns emerging from the body of scientific publications, using quantitative indicators to reveal underlying trends, relationships, and structures.

Clearly defining the object of study is the first crucial step in determining a bibliometric project. This involves precisely specifying the documentary corpus to be analyzed, the time periods considered, the data sources used, and the limitations inherent in these methodological decisions. A well-defined bibliometric study is characterized by research questions that cannot be answered by simply reading the literature, but instead require the systematic analysis of large volumes of bibliographic data to reveal patterns that transcend individual subjective experience. For example, while a researcher may have intuitions about trends in their field, bibliometrics allows them to quantify these trends, identify influential factors, and discover relationships that are not obvious at first glance.

The elements that distinguish a bibliometric study from other types of articles include the systematic use of quantitative indicators, the analysis of relational patterns using network techniques, the identification of temporal trajectories, and the contextualization of findings within theoretical frameworks on scientific dynamics.

A project becomes bibliometric when its main objective is to understand the architecture of knowledge, collaboration networks, citation patterns, or thematic evolution, rather than advancing substantive expertise within a particular field.

This specific methodological orientation determines both the design of the study and the structure of the resulting manuscript, requiring detailed methodological sections on data collection and processing, as well as a presentation of results that balances quantitative rigor with meaningful substantive interpretation.

Bibliometric research addresses knowledge gaps; that is, it serves a cognitive purpose, so the scientific problem will always revolve around a lack of knowledge about research trends. This is extremely important, as it should be used to set the objective and guide the entire article. All of the above is implicit in the INTRODUCTION section of the article.

10.2. Writing the methods

The METHODS section in a bibliometric study should provide a comprehensive description that allows for complete replication of the research. This methodological transparency is a fundamental pillar of scientific integrity in bibliometrics, where seemingly minor decisions during data collection and processing can significantly influence the results. The writing must balance the technical detail necessary for reproducibility with the clarity of presentation that facilitates understanding by readers from different fields. Well-documented methods not only validate the study's credibility but also advance the discipline's methodology by enabling comparison and iterative improvement of analysis techniques.

10.2.1. PRISMA diagram for transparency

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram (https://www.prisma-statement.org/) systematizes the document selection process through four sequential phases: identification, screening, eligibility, and inclusion. In bibliometrics, the identification phase documents the databases consulted, the search strategies used, and the records obtained, including duplicates. Screening applies criteria based on metadata such as document type, language, or accessibility.

Eligibility evaluates full texts based on their thematic relevance and metadata quality. The final phase specifies the corpus analyzed and the exclusion criteria applied. This structure ensures transparency in the construction of the bibliographic dataset, which is fundamental to the study's validity.

The bibliometric adaptation of PRISMA incorporates specific elements such as metadata extraction and cleaning protocols. This includes processes for author normalization, affiliation standardization, and terminology unification.

It should also document algorithms for resolving ambiguities in institutional names and managing duplicate records between databases. These expanded elements are crucial to ensuring the consistency of the analytical corpus. The flow visualization should reflect these additional steps, showing how raw data is transformed into the final dataset ready for bibliometric analysis.

Technical reproducibility completes the PRISMA framework through comprehensive documentation of tools and parameters. This includes bibliometric software versions, specific analysis configurations, and processing scripts. The thresholds applied in network analysis, the clustering algorithms used, and the normalization criteria for indicators must be specified in detail. The availability of code and datasets in accessible repositories enables independent verification of results. This methodological transparency enables exact replication of the study and facilitates comparisons between similar bibliometric research.

10.2.2. Ethical protocols

Ethical aspects of bibliometrics include considerations of data scraping, the use of sensitive information, and responsibility for interpreting and communicating results. The text should explicitly state compliance with the terms of service of the databases used, the measures implemented to avoid server overload during data extraction, and the protocols followed to ensure the secure storage and responsible use of the information obtained.

When analyzing individual researcher data, procedures for preserving anonymity and avoiding potentially harmful uses should be described, especially in evaluative contexts where results could affect career trajectories.

Disclosure of conflicts of interest and funding sources completes the fundamental ethical issues that should be addressed in the methods section.

The methodology of bibliometric studies can also be supported by a flow chart of articles, specifying the number of articles found and, from there, the flow of filters or selection criteria until the final number or sample to be worked with is obtained. Each stage of the flow should record the number of documents included and excluded, along with the specific criteria applied in each filter, allowing for immediate visualization of how the initial sample is progressively reduced to form the final bibliographic corpus for analysis. This is recommended for studies that do not work with all of the articles found, i.e., those that apply sampling, which must be declared.

10.3. Writing up the results

The presentation of bibliometric results should follow a logical sequence that guides the reader from general patterns to specific analyses. For this reason, it is always advisable to begin with fundamental productivity indicators: total number of publications, temporal distribution, and annual growth rates, as well as complementary criteria such as article type, countries, etc. These fundamental data provide the essential quantitative context for interpreting subsequent analyses. These results can be presented in tables or graphs showing the temporal evolution, highlighting periods of acceleration or deceleration in research activity.

The analysis continues with the most productive authors and institutions, using tables that rank the main contributors according to productivity and impact metrics.

This includes the number of publications per author, the institutions with the highest output, and the most active countries. These tables should be presented with well-defined columns showing ranks, names, number of documents, and total citations.

Impact and citation indicators make up the third section of results, which presents citation distributions, h-indices, and normalized indicators by field. Bar charts show citations per document, while boxplots visualize citation distributions and percentile tables establish normalized comparisons. These visual elements highlight the most influential publications and predominant citation patterns, requiring special attention to scales and ranges to avoid distortions in interpretation.

Network and collaboration analysis represents a critical visual component presented through annotated graph figures.

The sequence begins with co-authorship networks that show the main collaborative clusters, followed by co-word maps that reveal the field's thematic structure. Each figure includes detailed legends explaining the meaning of colors, node sizes, and link thicknesses, maintaining sufficient resolution to ensure readability even in article-size prints.

Density and heat maps are reserved for showing spatiotemporal patterns and thematic concentrations.

These visual elements are particularly effective for communicating the diachronic evolution of themes or the geographical distribution of scientific production. The design ensures color palettes are intuitive and accessible, using gradients that progress from low to high concentrations in a logical, consistently interpretable manner.

The application of Sankey diagrams reveals thematic mobility, changes in collaboration patterns, or the evolution of research foci.

The design of these diagrams maintains sufficient space between flows to ensure readability and uses distinctive colors for different trajectories. The complexity of these graphs requires detailed explanation in captions and references in the main text.

Temporal trend analyses are presented in a separate section, where time series of emerging concepts or evolutionary indicators are shown.

Line graphs show thematic trajectories while area graphs visualize the relative prominence of different topics over time. These elements include significant reference points, such as scientific policy launches or technological milestones, that contextualize the observed changes.

The final section presents specialized results such as burst analysis, citation patterns, or interdisciplinarity studies. These complex visualizations are reserved for the end, once the fundamental patterns have been understood.

Their inclusion as complementary figures deepens specific aspects of the study, ensuring that each advanced visualization maintains a clear justification in the research objectives.

Sometimes, results are not very relevant to the objective set out at the outset, so they should be dispensed with to give priority to those that really add weight to the research. Likewise, the order may change depending on the flow of information to be followed, ensuring that it progresses from the most general and fundamental to the most specific and advanced.

The numbering of tables and figures strictly follows the order of appearance in the text, with cross-references that guide the reader smoothly between different visual elements. Each table and figure is designed for independent comprehension, with descriptive titles and self-explanatory captions. This care for visual presentation distinguishes professional bibliometric studies from mere quantitative exercises, thereby enhancing the communicative quality of the academic manuscript.

10.3.1. Selecting visualizations by audience

The choice of visualizations in bibliometric studies must be specifically tailored to the technical knowledge and interests of the target audience.

For evaluation committees and scientific managers, priority is given to basic indicator graphs such as the temporal evolution of publications and citations, simplified collaboration network diagrams, and institutional productivity heat maps. These visualizations efficiently communicate essential information on productivity, impact, and collaborations without specialized knowledge of bibliometrics, facilitating decision-making based on robust, accessible quantitative evidence.

For academic audiences specializing in the field of study, visualizations can incorporate greater technical complexity and specific terminology. Detailed co-word maps, co-citation analyses, and thematic evolution diagrams are appropriate for these contexts. Presentations for specialists may include advanced methodological parameters, specific clustering algorithms, and network centrality measures, provided that these technical elements add substantive interpretive value to the study's findings.

When the bibliometric study is aimed at editorial boards or interdisciplinary audiences, visualizations should maximize immediate clarity and communicative impact. These visualizations transform complex data into intuitive narratives of the research field's dynamics, facilitating the transfer of bibliometric knowledge into contexts of science policy and strategic planning.

10.3.2. Tables for comparisons

Tables are the optimal format for presenting systematic comparisons between bibliometric entities such as authors, institutions, journals, or countries.

The tabular design should organize information according to clear sorting criteria, typically by productivity metrics or descending impact. Each table should include columns for ranking, entity name, number of publications, total citations, and normalized indicators, such as the h-index or average impact per document, enabling immediate multidimensional comparisons.

The hierarchical structure of tables facilitates the progressive analysis of comparative results. The initial tables present general productivity rankings, while subsequent tables show breakdowns by time periods, subject specialties, or types of collaboration.

This organization allows for the identification not only of the most productive entities globally, but also patterns of specialization, differential growth trajectories, and changes in relative leadership over time. Each table should be accompanied by brief textual analyses highlighting the most relevant comparative findings.

10.4. Writing the discussions

The discussion section is the central analytical component where bibliometric findings acquire substantive meaning through contextualized interpretation. This transition from quantitative description to qualitative explanation requires the systematic integration of empirical evidence with existing theoretical frameworks on scientific dynamics.

The construction of interpretive arguments is based on methodical triangulation among one's own findings, specialized literature, and scientific theories, establishing meaningful connections between observed patterns and underlying processes within the research ecosystem. The discussion should avoid both the mere repetition of results and unfounded speculation, balancing interpretive rigor with disciplinary relevance.

The articulation between results and previous literature represents a methodological process that requires specific strategies for documentary comparison. The systematic conceptual mapping technique identifies key publications that address similar phenomena in comparable contexts, establishing critical dialogues with previous studies.

The selection of literature for comparison should include previous bibliometric research on the same subject area, qualitative studies that explore dimensions not captured by quantitative metrics, and theoretical works that provide explanatory frameworks for interpreting the patterns observed. This documentary triangulation substantially enriches the interpretive depth of the discussion.

The identification of consistencies and inconsistencies with existing literature should be carried out through structured comparative analysis.

When the results confirm previous findings, the discussion should explore the underlying

mechanisms that would explain these transcontextual regularities. In light of discrepancies with prior studies, the analysis should examine methodological factors, temporal differences, or contextual particularities that could account for the observed variations. This comparative approach transforms mere coincidences or differences into opportunities to advance theoretical understanding of the bibliometric phenomena under analysis.

10.4.1. Link to scientific policies

The discussion of implications for scientific policies is an essential component that connects bibliometric analysis with practical applications in research management. The interpretation of results should explain how the observed patterns inform the design, implementation, or evaluation of interventions within the scientific system.

For example, identifying underrepresented subject areas can support recommendations for specific funding programs, while analyzing collaboration networks can inform internationalization strategies. This link between bibliometric evidence and policy action adds substantive relevance to the study.

Developing policy recommendations requires a careful balance between empirical evidence and contextual considerations. Proposals should be logically derived from the results, avoiding undue extrapolations beyond what the data allow.

The discussion should explicitly acknowledge the partial and complementary nature of bibliometric evidence within complex policy-making processes. This rigorous approach strengthens the study's credibility and its potential to improve research systems.

The identification of plausible causal mechanisms linking policy interventions to observed bibliometric patterns significantly enriches the discussion.

For example, changes in international collaboration patterns may be related to specific scientific mobility programs, while variations in thematic productivity may be associated with established funding priorities. This theoretical elaboration transcends mere correlation to propose well-founded causal explanations that inform future science policy interventions.

Contextualizing findings within existing science policy frameworks adds analytical depth to the discussion. Contrasting them with national and international science planning documents, reports from organizations such as UNESCO or the OECD, and evaluations of specific funding programs allows the results to be placed within broader political trends.

This contextualization facilitates the identification of mismatches between stated science policy objectives and actual patterns observed in research activity.

The development of prospective scenarios based on identified bibliometric trends represents a particularly valuable discursive strategy for scientific planning. The projection of thematic trajectories, the identification of emerging opportunities for specialization, and the mapping of future scientific capabilities transform the retrospective analysis typical of bibliometrics into a tool for strategic anticipation. These prospective exercises must be clearly grounded in the data presented and acknowledge their predictive limitations.

10.4.2. Methodological limitations

The exhaustive recognition of methodological limitations constitutes an element of

intellectual rigor in the writing of bibliometric discussions.

The partial coverage of databases represents a fundamental limitation that should be characterized quantitatively when possible, specifying the geographical, disciplinary, and typological biases introduced by the selection of sources.

The discussion should assess how these limitations affect the external validity of the findings and appropriately delimit the scope of the conclusions. This methodological transparency enhances the study's credibility.

Methodological decisions in data processing introduce additional limitations that deserve critical examination in the discussion. Author disambiguation algorithms, terminology normalization criteria, and thresholds in network analysis are potential sources of bias that can affect results. The discussion should explore how different methodological decisions could have altered the patterns observed, demonstrating awareness of the constructed nature of bibliometric data. This methodological reflexivity characterizes high-quality bibliometric research.

The identification of limitations should be complemented by proposals for future research directions that overcome current methodological constraints.

The discussion may suggest strategies for validating findings through triangulation with other data sources, alternative methods of analysis, or replication across different temporal or disciplinary contexts. This projection toward future methodological advances positions the study within an evolving research program rather than as an isolated exercise.

The analysis of the differential impact of methodological limitations on different types of findings adds sophistication to the critical examination.

Some limitations mainly affect quantitative productivity analyses, while others particularly distort relational structures in network analyses. This differentiated characterization allows readers to assess which findings are more methodologically robust and which should be interpreted with greater contextual caution.

10.4.3. Integration of multidisciplinary perspectives

The incorporation of theoretical perspectives from different disciplines substantially enriches the interpretation of bibliometric findings. Conceptual frameworks from the sociology of science, the economics of innovation, and science, technology, and society studies provide valuable interpretive lenses for understanding the patterns observed. This multidisciplinary integration transcends mere bibliometric description to advance toward substantive explanations of the social and institutional dynamics underlying the quantitative data.

Triangulation with qualitative evidence on the phenomena under study is a compelling strategy for enriching the bibliometric discussion.

The incorporation of interviews with researchers, analysis of science policy documents, or ethnographic studies of laboratories, when available, allows quantitative patterns to be contextualized within specific social processes. This methodological integration overcomes the inherent limitations of purely quantitative approaches in bibliometrics.

Examining the epistemological implications of bibliometric findings adds theoretical depth to the discussion. Analyzing how the observed patterns reflect or challenge established conceptions of knowledge production, the structure of scientific disciplines, or the dynamics of paradigmatic change connects bibliometrics to fundamental debates in the philosophy of science. This elevation of intellectual dialogue maximizes the theoretical contribution of the study.

10.5. Good practices in data visualization

Adequate visualization of bibliometric data requires the systematic application of graphic design principles that balance analytical accuracy with communicative clarity. Color selection should prioritize perceptually uniform palettes and ensure accessibility for people with color blindness through specific verification tools.

Color schemes should be applied consistently across all visualizations in the study, using intuitive conventions: warm colors for high values and cool colors for low values. This visual consistency facilitates comparative interpretation between different figures and graphs.

Visual hierarchy is a fundamental principle for guiding the reader's attention to the most significant elements of each visualization.

The strategic use of size, contrast, and position highlights key patterns without distorting the integrity of the underlying data. Text labels should be carefully placed to maximize readability without obscuring important information, and connectors should be used when necessary to associate text with specific graphic elements. This hierarchical organization transforms complex visualizations into intuitive visual narratives.

The choice of visualization format should be based on the nature of the data and the specific communication objectives. Bar charts are ideal for comparing magnitudes between discrete categories, while line charts effectively show temporal trends. Network visualizations capture structural relationships, and heat maps represent densities or intensities between continuous dimensions. Each formative selection should be justified by its ability to communicate the analysis's central findings efficiently.

The scalability of visualizations warrants special consideration in academic publications. Figures should be designed to maintain legibility in both print and digital formats, using sufficient resolutions to allow zooming without loss of quality.

The balance between level of detail and visual clarity is optimized through progressive simplification techniques that present general information initially, allowing exploration of details through complementary or interactive visualizations.

Contextualizing visualizations through appropriate reference elements substantially improves their interpretability. Clearly marked scales, significant reference lines, and explanatory annotations situate the data within relevant interpretive frameworks.

This contextualization should include indicators of statistical significance when applicable, as well as comparisons with disciplinary averages or reference values established in the specialized literature.

Methodological transparency in visualization requires complete documentation of all transformations applied to the raw data. Smoothing, aggregation, or normalization procedures

should be explained in detail, allowing readers to understand how processing decisions affect the patterns visualized. This transparency extends to the declaration of specific software and parameters used to generate each visualization, facilitating replication and independent verification.

Iterative usability testing with target audience representatives identifies opportunities for improvement in the visualizations. This process of continuous refinement ensures that the visual elements effectively communicate the findings to readers with varying levels of familiarity with bibliometric techniques. Incorporating feedback on aspects such as color interpretation, symbol comprehension, and interactive visualization navigation optimizes the final communicative effectiveness.

Universal accessibility represents a fundamental ethical principle in bibliometric data visualization. The design must consider diverse needs by providing alternative textual descriptions, ensuring sufficient color contrast, and avoiding exclusive reliance on color to convey critical information. These accessibility considerations broaden the study's reach and impact and align bibliometric practice with contemporary standards of academic inclusivity.

10.6. Other sections in bibliometric studies

Bibliometric studies may incorporate various additional sections depending on the specific requirements of each academic journal. Conclusions are the most common component, summarizing the main findings and their relevance to the field of study. This section should offer a concise overview that goes beyond a mere repetition of results, integrating the theoretical, methodological, and practical implications identified in the discussion. Effective conclusions establish clear connections between the study's initial objectives and its achievements, and outline meaningful directions for future research.

Acknowledgments are another common section that recognizes contributions not sufficient for authorship and sources of funding. This section should clearly specify the role of each collaborator mentioned and declare any potential conflicts of interest. Acknowledging technical support, expert advice, or access to research infrastructure enhances the study's academic transparency. Funding sources are detailed using persistent identifiers when available, facilitating the traceability of institutional support.

Appendices and supplementary material complement the main narrative without interrupting its flow. This section may include detailed methodologies, processing algorithms, additional visualizations, or extensive datasets. The organization of supplementary material should follow a logical structure with explicit cross-references from the main text. Each item included justifies its presence through its value for replication or further analysis of the presented findings.

Some specialized journals require specific sections such as practical implications, graphic summaries, or data availability statements. Practical implications translate bibliometric findings into concrete recommendations for different stakeholders in the scientific system. Graphic abstracts visually synthesize the study's main contributions, facilitating its dissemination across academic and professional channels. Data availability statements detail conditions of access to the datasets used, aligning with the principles of open science.

The standardization of these additional sections varies significantly across disciplines and academic journals. Careful review of the guidelines for authors before manuscript submission ensures compliance with specific requirements. Adapting the study structure to these editorial

requirements optimizes its evaluation during peer review. The next chapter will examine in depth the criteria for selecting target journals and strategies for navigating the academic publication process in bibliometrics.

Recap

- The IMRyD structure improves clarity and consistency.
- Methods should include sources, filters, and collection dates.
- Publishing data and code promotes transparency.
- Figures should be self-explanatory and reproducible.
- The abstract should consist of the objective, method, and key results.
- Supplementary tables keep the body of the text organized.
- Standardization should be explained.
- Acknowledging limitations increases credibility.
- Discussion should relate findings to previous literature.
- R Markdown and Jupyter integrate text and analysis.
- Supplementary material increases reproducibility.
- Clarity, precision, and concise style are essential.
- Avoid overinterpreting correlations.
- Follow each journal's guidelines.
- Respond to reviews professionally.
- Declare conflicts of interest.
- Include ORCID and standardized affiliations.
- Acknowledge technical contributions.
- Adapt language according to the audience.
- Organize work files systematically.

Self-assessment questions

1. What sections make up the IMRyD structure?
2. What should the Methods section contain?
3. Why publish data and code?
4. What characterizes a self-explanatory figure?
5. How should an effective abstract be structured?
6. How should limitations be described in a manuscript?
7. What advantages does RMarkdown offer?
8. How important is it to declare conflicts?
9. How can you improve the visibility of your article?
10. What should be included in supplementary material?

BIBLIOGRAPHY

1. Schimel J. Writing Science: How to Write Papers That Get Cited and Proposals That Get Funded. Oxford: Oxford University Press; 2012. ISBN: 9780199760237.

2. Alley M. The Craft of Scientific Writing. 4th ed. New York (NY): Springer; 2018. ISBN: 9783319695437.

3. Day RA, Gastel B. How to Write and Publish a Scientific Paper. 6th ed. Santa Barbara (CA): Greenwood; 2012. ISBN: 9781619250002.

4. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. Chichester: Wiley; 2009. ISBN: 9780470057247.