Aug 27, 2020

Workflow for retrieving all the data of the analysis introduced in the article "Citing and referencing habits in Medicine and Social Sciences journals in 2019"

DOI
dx.doi.org/10.17504/protocols.io.bbifikbn

Erika EAS Alves dos Santos[1], Silvio Peroni[2], Marcos Luiz Mucheroni[3]

[1]Fundacentro / Brazilian Ministry of Economy;
[2]Digital Humanities Advanced Research Centre (DHARC), Department of Classical Philology and Italian Studies, University of Bologna, Italy / Research Centre for Open Scholarly Metadata, Department of Classical Philology and Italian Studies, University of Bologna, Italy;
[3]School of Communication and Arts (ECA), Department of Information & Culture (CBD), University of São Paulo (USP), Brazil

Erika EAS Alves dos Santos
Fundacentro / Brazilian Ministry of Economy, School of Commu...

DOI: dx.doi.org/10.17504/protocols.io.bbifikbn

External link: https://www.scimagojr.com/

Protocol Citation: Erika EAS Alves dos Santos, Silvio Peroni, Marcos Luiz Mucheroni 2020. Workflow for retrieving all the data of the analysis introduced in the article "Citing and referencing habits in Medicine and Social Sciences journals in 2019". **protocols.io https://dx.doi.org/10.17504/protocols.io.bbifikbn**

Protocol status: Working
We use this protocol in our workspace, and it is working

Created: January 18, 2020

Last Modified: August 27, 2020

Protocol Integer ID: 32039

Keywords: SCImago Journal and Country Rank, Sampling, Articles sample, Journals sample,

Abstract

This protocol stablishes a method for selecting journals and articles for composing a research sample based on SCImago Journal and Country Rank – SJR metadata (available from https://www.scimagojr.com/). It might be applied partially or in full, in order to obtain an stratified and representative sample of journals and/or articles, considering a multidisciplinary, as well as a unique discipline approach, according to the main objective(s) of the research in which it will be applied.

Materials

For delimiting the journal titles that will compose the universe of the research it will be considered the metadata from SCImago Journal and Country Rank – SJR (available from https://www.scimagojr.com/).

SCImago is a public database that compiles metrics information from the journals indexed by Scopus database, which was launched in 2004 and is one of the most authoritative sources of citation data (SHOTTON, 2018) where journals are assigned to 27 major thematic categories as well as to 313 specific subject categories according to Scopus database classification.

The selection of this database is supported and justified by the expressive numbers of journals titles ranked (considering data from November 2019) and also by the detailed analysis of the metrics data for each journal title, what includes the index of citations received by each journal in the previous 4, 3 and 2 years.

SHOTTON, D. Funders should mandate open citations. **Nature**, London, v. 553, p. 129, Jan. 2018. Available at:https://www.nature.com/magazine-assets/d41586-018-00104-7/d41586-018-00104-7.pdf. Access in: Nov. 11th, 2019.

Before start

Journals and articles are usually considered for composing the universe of researches in different areas of knowledge aiming most diverse ends, among which state-of-the-art researches can be cited as one of the most common.

In the scientific scenery in which registered information is exponentially growing, the representation of information also takes on a rothele of linking researchers and information, guided by precision and agility in information retrieval. Being part of the universe of descriptive representation, bibliographic references usually act as instruments that first establish connections between information and its reader.

This approaches leads towards a reflection on what is, and what will be the role of bibliographic references in the citation network scenario, considering this as a facet of descriptive representation, which is passing through a huge conceptual restructuring, and also, what are the possible impacts of these changes in bibliographic metadata representation and citation network.

The main objective of this protocol is to propose a method for guidance and support for investigations in order to answer the following questions:

1. Considering the introduction of technological tools such as reference manager software and the instructions provided in the reference styles adopted by journals, were they to address fully all the problems pointed out by the study made by Sweetland in 1989 (https://doi.org/10.1086/602160)?
2. Which are the main possible causes for errors in mentioning, quoting and referencing practices?
3. Do bibliographic references and in-text reference pointers refer to the same cited entity at a conceptual level in terms of the FRBR specification?
4. What is the potential impact of using FRBR within bibliographic catalogues on information retrieval?

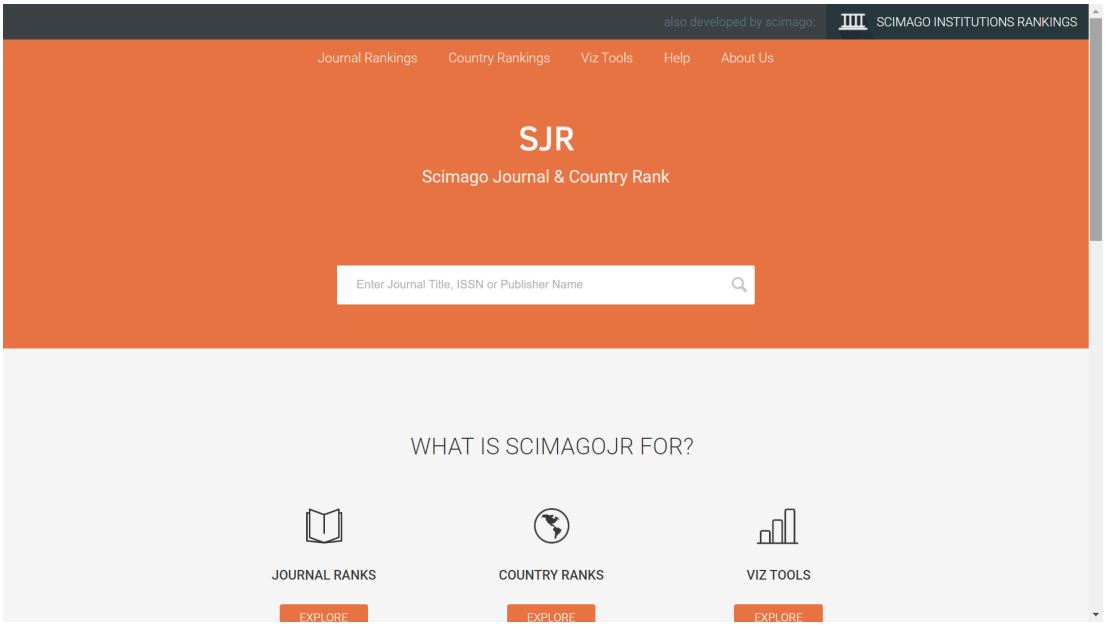
Since bibliographic references are elements of interest for all subject areas, this Protocol is based on a wide and multidisciplinary approach considering the subject areas arrangement proposed by SCImago database and, proposes a method for selecting journals and articles for composing research samples and analyzing such data, focusing on the aforementioned key questions.

The protocol aims to be comprehensive enough to be applicable, partially or in full, in diverse areas of knowledge, considering the diverse possible purposes, comprising the 27 areas of knowledge in which it is subdivided: Medicine, Social Sciences, Arts and Humanities, Engineering, Agricultural and Biological Sciences, Biochemistry, Genetics and Molecular Biology, Computer Science, Mathematics, Environmental Science, Business, Management and Accounting, Psychology, Materials Science, Earth and Planetary Sciences, Physics and Astronomy, Economics, Econometrics and Finance, Chemistry, Pharmacology Toxicology and Pharmaceutics, Nursing, Chemical Engineering, Neuroscience, Immunology and Microbiology, Health Professions, Energy, Decision Sciences, Veterinary, Dentistry, and Multidisciplinary.

Obtaining SCImago database journals title list

1 First, a search should be conducted in order to retrieve the titles of the journals indexed by SCImago database, according to the description in figure 1:

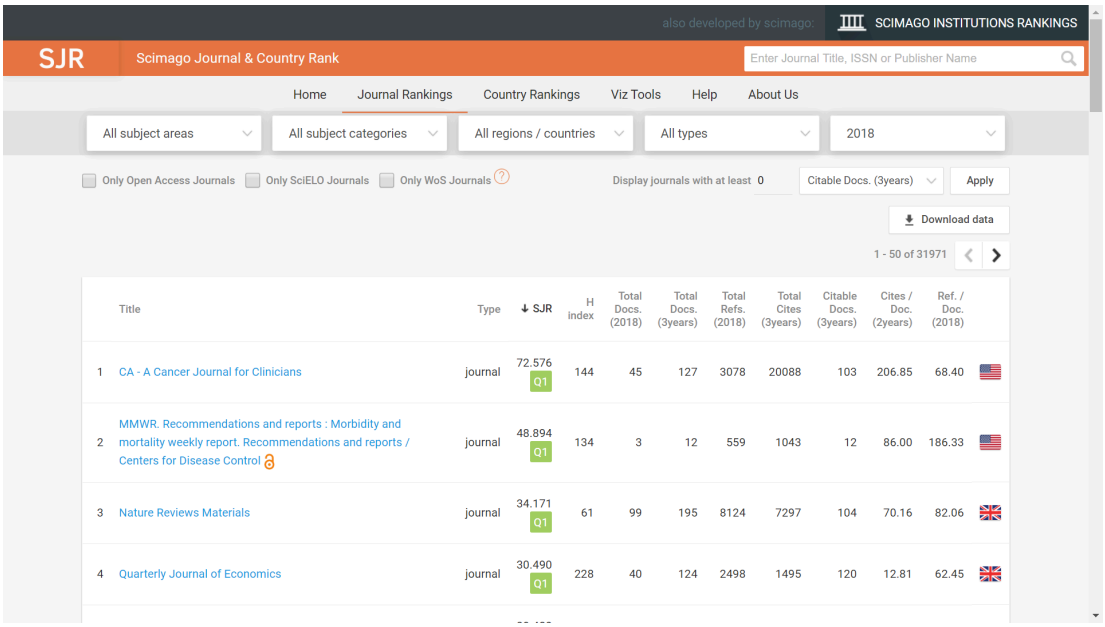
Figure 1 – SCImago database homepage



Source: **https://www.scimagojr.com/**

By clicking at the “Journal Rankings” hyperlink, the user will be led to the rankings search page (figure 2):

Figure 2 - SCImago database searching page



Source: **https://www.scimagojr.com/**

In response to this command, the database will return the general journal & country rank, from where it should be applied the delimitation parameters.

Delimiting parameters and filters for SCImago journals rank search

2 This step should be repeated as many times as the subject areas where journals should be selected. For each new search, the corresponding subject area of interest should be selected in the respective options box.
For the following sub-steps, please consider the Figure 2 - SCImago database searching page, given in step 1.

2.1 Select the subject area of interest in the "All subject areas" selection box.
The subject category is a criterion that should be considered in a whole at this moment. So, the “all subject categories” option, selected by default, should be maintained.

2.2 Select the country or the countries corresponding to the nationality of selected journals.
In case of the nationality of the journal is not a point of interest for the search, leave the "all regions/countries" checkbox filled in with "All regions/country" as default.

2.3 The SCImago database ranks several types of publications: journals, book series, conferences and proceedings and also trade journals.
Since the purposes of this method are to select journal titles, the “journals” option should be selected in the "All types" selection box. Nevertheless, the same criteria might be applied to select the other types of publications indexed by SCImago database. In this case, replace the "journal aspects" by the corresponding type of material selected in the next steps, eventually carrying out necessary adaptations, where applicable, since journals have particularities

regarding the periodicity of publication, which are not necessarily shared with other types of publications.

2.4 Select the year of publication to which the search results should refer.

2.5 Right below the selection boxes, there are three checkboxes, which will delimit the search results while the modality of access and the indexing source.
The first checkbox, "only open access journal", if checked, should restrict the search to records of open access journals. Leaving it unfilled will expand the search results to the whole records database, regardless the modality of access of the publications described in each record.
The second checkbox, "Only Scielo Journals" should restrict the search to the journals indexed by The Scientific Electronic Library Online - SciELO, which is "an electronic library covering a selected collection of Brazilian scientific journals. [...] The Project envisages the development of a common methodology for the preparation, storage, dissemination, and evaluation of scientific literature in electronic format." (SCIELO, 2019). Leave this checkbox unfilled to consider both records of publications indexed by Scielo as well as records of other publications that integrate the SCImago database in the search results.
By checking the third checkbox "Only WoS Journals", the search will restrict the results only to journals indexed by the Web of Science (WoS) database, similar to the process described for the previous checkbox.

SCIELO. Available at: https://www.scielo.br/scielo.php?script=sci_home&lng=en&nrm=iso. 2019. Accessed May 3rd, 2019.

2.6 On the right of the three checkboxes described in section 2.5, there is an editable field "display journals with at last ____", followed by a selection box. These two tools allow the researcher to delimit the minimum number of citable documents [option "citable docs. (3 years)" in the selection box] or the minimum total number of citations received by the journal titles retrieved by search results [option "total cites (3 years)" in the selection box]. Whenever applicable, manually fill the editable field with numeric characters and select the option from the selection box which better corresponds to the search aims.

2.7 Click the button "apply" button to override the delimited parameters on the search results.

2.8 Tip: for optimizing the data manipulation, it is suggested to download the search results to an excel sheet, by clicking the button "download data".

Selecting the journal titles to compose the universe of the research

3 As described in step 2.3, this protocol will refer specifically to journal selection, considering especially the particular journal aspects regarding periodicity, from this step on. For

applying this method on the selection of other types of publications rather than journals, adaptations should be carried out from this point on.
SCImago database comprises a wide coverage of journal rankings and there might be noticed discrepant amounts of journal titles comprised of each subject area. In addition, each journal volume usually comprises plenty of issues published by year, and these indices also might be very different among different journal titles. That makes it impracticable to consider the total number of journal titles in its completeness, and that justifies the elaboration of a method for selecting a representative sample of the total amount of journal titles.
It is proposed the establishment of a sample that maintains the same proportions of representativeness in relation to the total volume of journal titles indexed by SCImago, considering each area of knowledge individually. That is to say that the steps described from 2.1 up to 2.7 should be repeated for each subject area comprised of the study for which the journal selection is being done.

3.1 Verify the total number of journals indexed under the parameters established for the research. This number corresponds to the findings achieved by the execution of step 2 and is the one exhibited right below the button "download data".

3.2 Proceed the search described in step 2 for each subject area of interest to the study. Take note of the total number of journal articles indexed in each subject area of interest, and then, establish the representativeness of this subject area over the total volume of journal indexed in the SCImago database. For this, divide the total number of journal titles indexed under a specific subject area, by the total amount of journals indexed in the SCImago database as a whole (steps 2 and 3.1). Then, multiply this number by 100. The result of this operation will be the percentage that such area of knowledge corresponds in the total of titles of indexed journals, that is, the percentual index of representativeness of each specific subject area in theSCImago database. That is why this process should be repeated for each subject area. The index of representativeness will determine the percentual portion of journals about to be selected for each subject area.

3.3 For the cases in that the percentual volume of journals to be select results in a decimal number in the total number of journals indexed under a specific subject area, this number should be rounded according to mathematical principles.
The minimum number of journals admitted for each subject area is 2, regardless of the percentage equivalence or the rounding result in the cases it is applicable.

3.4 Once defined the number of journals that should compose the sample, the procedures to select the journal titles within each discipline should be carried out. Journals titles occupying the highest positions in the SCImago ranking should be the ones selected for the sampling, obeying the following three criteria, which are presented in specific columns in the search results page in the SCImago database (or in the excel sheet, as mentioned in step 2.8):

1. First criteria: the total cites ranking. The sample should consider the most cited journals of each subject area. This metric should be considered as a proxy for the prestige of journals,

assuming that the most citations one journal receive, the most relevant it is supposed to be in their particular subject areas.

2. Second criteria: the publisher. There should not be admitted more than one journal of the same publisher in the same subject area. This will assure a heterogeneous sample and the coverage of the different biases adopted by the editors regarding the presentation and formatting of citations and references in scientific articles.
3. Third criteria: the subject category. SCImago database classifies journal titles by subject areas and further by subject categories, defining specific subheadings for particular journals. Some journals may be classified simultaneously in more than one subject area and/or category. To avoid possible overlaps between the selected journals in terms of areas and categories, the sample should consider only journals classified under a single subject area and subject category.

The sample will consider the best-positioned journal titles in the total cites Scimago ranking. In case of any journal selected in this preliminary selection does not comply to any of the previous selection criteria, this (or these) journals should be replaced by the immediately subsequent journal title in the “total cites” ranking. This process must be repeated until the target number of selected journals title determined for each subject area is reached.

A brief example of the application of steps 1 to 3 for reference may be found here.

Example.pdf

Selecting journals sampling chronological parameters

4 Some journals do not have a regular periodicity, e.g., some electronic journals publish articles insofar as they are being approved and accepted for publication by the editorial board. Some of them do not even mention a specific issue number for each journal issue. Facing this, the eligibility criteria adopted for choosing issues from the selected journals should establish a chronological parameter. That is to say that the chronological coverage of the sample should be in line with the scope of the study, be it a specific month, semester, year, etc.

Due to the huge volume of data released by publishers of periodical contents, it is suggested to choose a relatively short chronological interval, such as a specific month, in order to obtain a volume of data which is massive enough to support a scientific study, but not making it too big to make data manipulation unfeasible.

Selecting the articles to compose the universe of the research

5 For each journal issue, it should be selected 5 articles, assuming that articles from the same journal are subject to the same editorial policies, and then, the variance of data collected from articles published by the same journal title tends not to add significant data to the analysis discussion. So the analysis should prioritize as many journals as possible

per knowledge area, rather than increasing the number of articles analyzed per issue, in order to ensure a wider and more comprehensive coverage of scientific articles from different disciplines.

The procedures for the selection of the articles are based on a random and exponential method of selection, as exemplified in the schema represented below, where the "X" means the average of articles published by the selected journals in the chronological period defined for the sampling on step 4:

$$X^{th}; 2X^{th}, 3X^{th}, 4X^{th}, 5X^{th}$$

So, for each journal issue, the first selected article will be the one occupying the "X^{th} position" in the issue summary, the second should be the one occupying the position $2X^{th}$ and so on, successively until the 5th article is selected.

For example, considering that 30 is the average of articles published by selected journals in chronological period of coverage of the research, the 30^{th}, 60^{th}, 90^{th}, 120^{th}, and the 150^{th} articles of each issue will be the selected ones.

For proceeding this step it is important to keep the three following instructions in mind:

1. For the cases in which the number of articles published within a selected journal issue is smaller than the average number of articles published by the journal in the chronological period defined in step 4, consider a cycle and systematic counting process. This ascending counting should start from the first article appearing in the summary of the selected journal issue. Once the counter reaches the last article of the summary, consider a return to the first article of the summary and continue the counter from there. This process should be repeated until the 5th article is selected.
2. When an article is chosen, the counting should be reset and restarted from the article immediately following the selected one.
3. The selected articles should be disregarded by the following countings.

Some notes to this step:

1. Only articles introducing original research communications should be considered for sampling. Since the terminology applied by different publishers to name their respective journal sessions and define the type of articles they publish is not uniform and also, considering that not every journal designates original communications as so, the sample should consider the following terms as synonyms of original communications: articles, papers, researches, original papers, original articles, original researches, research papers, research articles, regular papers, regular articles, regular researches, and regular articles. Exceptions may be admitted whenever a selected article, clearly introducing one original research communication, is not classified under one of the previous terms by the journal in which it is published.
2. In case of a selected journal issue does not publish any original communications within the period defined in step 4, it should not be considered eligible for composing the sample and, therefore, should be replaced by the immediately previous eligible issue under the same journal title. Nevertheless, an eligible issue should not be discarded if there are less than 5

eligible articles, that is to say, that having at least one eligible article makes a journal eligible for the sample as well.

3. For researches in which the bibliographic references metadata correspond to the core research object, the sample should consider articles whose bibliographic references sessions are presented in an exclusive list, designated for the sole purpose of providing the bibliographic references of the cited works within the article, regardless of how this session is named by the publisher. Exceptionally, It should be admitted articles whose bibliographic references are distributed in footnotes throughout the text, provided that the format in which they are presented corresponds, or is similar to a widely accepted reference style such as Vancouver, Chicago, or APA. Similarly, journals in which articles bibliographic sessions are mixed up with text notes should be considered eligible for the sample only in cases in which it is possible to clearly differentiate notes from bibliographic references. Journals that do not meet these requirements, should not be considered eligible for the sample, even if they meet all the previous criteria.
4. Some journals divide their summaries into subject categories. In these cases, articles whose thematic is multidisciplinary might be simultaneously considered in more than one subject category. That is to say that it is true that one article may appear twice in the same summary, in different subject categories. For these cases, the counting should consider only the first appearing entry of the article title within the summary. All the following duplicated entries of a specific article should be excluded from the counting, in order to ensure that the likelihood of selection across the articles is balanced.
5. Articles not assigned to any issue should not be considered eligible for the sample. Exceptions may be admitted for the specific cases in which the publisher does not attribute numbers to its journal issues or, journals actually not publishing in issues. In such cases, all the articles published in the immediately previous period equivalent to that established in step 4 (e.g. the previous month or the previous semester), should be considered as a virtual regular issue.
6. Non-regular issues (e.g. supplements, special issues, thematic issues, commemorative issues, etc.) should not be considered eligible for sample composition purposes. In case the selected issue corresponds to one of these cases, it should be replaced by the immediately previous issue fully attending the selection criteria.
7. It should be prioritized for the .pdf version of the articles. Exceptionally, in cases that this format is not made available by the publisher, the sample might consider other available formats and versions, e.g. HTML format, or printed (hardcopy), not necessarily in this order or priority. If it is not possible to have access to the articles of a selected issue in any way, the journal should be replaced by the immediately following eligible journal.

Lastly, it is important to clarify that:

1. The average number of articles published by all selected journals in the period covered by the survey (step 4) should not be admitted as a criterion for dimensioning the size of the sample of articles. Since some journals publish hundreds of articles per month, the monthly average of published articles would increase considerably, and this might mask the real

panorama of articles publishing. Also, considering the average number of articles published by the selected journal as the determinant of the sample sizing might put some journals (especially the less representative ones) in a disadvantage position, since the number of articles published by some journals is much fewer than the monthly average of published articles in the period covered by the research. Considering these rates individually by journal would avoid these issues to influence the sample composition.

2. There are no established standards to determine the order of appearance of articles within journal issues. Some publishers structure the summary of their journals by subjects, others by date of acceptance, by authors, etc., and there are also publishers who do not even have an established criterion for that. So, by choosing the first 5 appearing articles of each selected issue, means admitting the risk of analyzing 5 articles under a specific approach, e.g. the same subject in some cases. This justifies the random method for selecting articles within issues, exposed in the previous items, and assures the heterogenicity of the sample.

The journal analysis

6 The journal analysis should consider general aspects of the journal and articles' presentation themselves and the way in which publishers provide information to potential authors regarding the rules for formatting mentions, quotations, bibliographic references, and instructions regarding submission processes. All data collected in the ambit of step 6 should consider a quantitative approach, that is, the results should be converted into numbers, representing the total of elements attending to the main specifications of each specific observation or, into percentual indexes. The following items describe the items that should be considered in the journal analysis level.

6.1 The modality of access: there should be verified whether the modality of access to the journal is open access or restricted access. Within the scope of this method, open access, which is considered a synonym of “freely available”, is understood by the unrestricted online access to a determined content, free of costs or other access requisites. At first, there are three accessing modalities for journals: freely available (i.e. works which can be accessed online, free of charge), restrict access (i.e. when it is required payment for accessing a specific content), and, the mixed access, (i.e. journals which makes available both open access content and restrict access content). Mixed access journals should be considered as not freely available journals. For identifying the modality of access of journals, it is important to clearly specify the used browser and version as well as the nationality of the IP of the computer used for the queries, since these two variables may alter the modality of access to the articles, while open access or restricted access.

6.2 The reference style adopted: the identification of the reference style adopted by each selected journal will allow the identification of the most used reference styles in each subject area. Also, it will be possible to define how publishers are susceptible to develop their own reference styles. The analysis should consider data provided by the instructions for authors, provided within the publishers' webpages.

6.3 The customization of reference styles: for those publishers adopting widely accepted reference styles like Vancouver, Chicago, AMA, and APA, the analysis should consider whether customizations are added to the adopted reference style guidelines according to the publisher's specific needs, i.e. suggest changes on the number of authors to be considered within in-text reference pointers and bibliographic references for works with multiple authorship in relation to the adopted reference style. This approach will allow the identification of the level of explicit influence of publishers on the way bibliographic references, mentions, and quotations are presented in their respective journals.

6.4 The maximum number of bibliographic references allowed per article: Some discussions demand a more in-depth scientific basis than others. In these cases, it is common that articles are plenty of mentions and quotations. However, considering the purposes and the general length of an article, it is not recommended to have too extended bibliographic references lists. First, because having plenty of quotations and mentions in a text might make its reading tiring and boring, in addition to obfuscating the author's own arguments. Second, because the discussion presented in an article should be not too short as not to include the exposition of purposes, methods, results, and the author's arguments, nor too long in order to be out of character and confused with other types of publications. Since there is no convention regarding the range of the length of bibliographic references lists (and in fact, there are no means for it, without impairing the specificity of each work that may require longer and in-depth discussions), authors are expected to use common sense when writing their scientific productions. On the other hand, some publishers use templates for diagraming articles and, in these cases, there might be some specific area for bibliographic references lists, which no possibilities of length adjustments and so, the author have to write their texts in order to fit in that specific space delimited by the publisher.

6.5 The way instructions for publishing are transmitted to authors: this topic is not about the instructions for publishing properly. This is regarding the way instructions for presenting and formatting mentions, quotations, and bibliographic references (the reference style itself) is presented to the authors. Some publishers indicate a specific reference style to be used in their publications, which can be widely accepted ones, i.e. Vancouver and Chicago or even self-authored reference styles. Others give few examples on how to present and format citation metadata, extracted from widely accepted reference styles, to support authors' decisions regarding presenting and formatting bibliographic metadata. usually, in these cases, there is any orientation regarding the way of presenting mentions or quotations, whether between single or double quotation marks, when indents should be applied, etc. Other publishers accept submissions formatted according to any reference style and make available a team that is responsible for the formatting of citation data for accepted works according to the journal's style (not always free of charge for the author). The approach referring to this item should consider the way these formatting instructions are provided to the authors by publishers and then evaluate the level of clarity of these guidelines, classifying them into clear or unclear. For instance, instructions providing an

explicitly clear and comprehensive set of instructions for managing citation data, or indicating a specific external widely accepted reference style, i.e. Vancouver reference style, should be classified as clear. On the contrary, guidelines not clearly indicating a widely accepted reference style or not providing clear and comprehensive instructions on presenting and formatting citation metadata, i.e. those not providing instructions on how to describe certain types of publications, like thesis, and conference papers, should be classified as unclear.

6.6 The recommended reference management softwares: There are plenty of reference management softwares available to the scientific community, some made available free of charge, some not. Some publishers recommend the use of these tools, some do not. In order to identify the level of the editorial appeal regarding the use of these instruments, the analysis should consider whether the publishers recommend the use of these tools or not. For the cases in which publishers make available a specific plugin for any reference style management software, or even mention a specific software, the analysis should consider a reference management recommendation.

6.7 The bibliographic reference lists titles: The way publishers name the bibliographic reference sessions within their journals is not uniform. To detect the range of this variation, the analysis should consider and evaluate the various ways in which the session of bibliographic references is referred to within articles.

6.8 The bibliographic references lists assortments: there are commonly two ways of bibliographic references assortment: the citation-sequence, i.e. when in-text reference pointers referring to mentions and quotations are specified using a number corresponding to a particular bibliographic reference in a bibliographic references list arranged in ascending numerical order, which is usually used by the medical journals and related areas, and the author-date assortment, i.e. when in-text reference pointers referring to mentions and quotations includes author's surnames and year of publication of the cited works described in a particular bibliographic reference, which is more commonly used by the humanities researchers. the analysis must quantify the rate of articles using each citation system.

The article analysis

7 The article analysis shall be more in-depth and detailed than the previous ones and comprises the most substantial part of data supporting the discussions. Similar to the previous steps, all data collected in the ambit of step 7 should consider a quantitative approach, that is, the results should be converted into numbers, representing the total of elements attending to the main specifications of each specific observation or, into percentual indexes.

7.1 The modality of access to the article: similar to what was considered by substep 6.1, this substep refers to the modality of access of the article, since restricted access journals also may provide freely available content. So approaches on this substep should consider the

description provided previously on substep 6.1, replacing the journals' perspective by the articles' perspective. on the same line, the browser and version as well as the nationality of the IP of the computer used for the queries should be considered, since the modality of access to the article can vary in accordance with the combination of these features. One important note to keep in mind during this analysis is that some journals composing the sample are indexed by PubMed Central. So, it may happen that a restricted access article can be freely accessed through this database. For the purpose of determining whether an article is freely available or restricted access, the analysis should consider the decision of the publisher, that is, the modality of access perceived for a specific article within its' publisher website or database.

7.2 The format of the article file: some publishers make their articles available in PDF. files, others in HTML, and others in both formats. Facing this variation, the analysis should quantify the sample of the articles, according to the format in which they are made available by their publishers.

7.3 The exportation of citation metadata: some publishers provide tools for exporting the citation metadata of the articles they publish within their journals. These tools may export these metadata both in a machine-readable format or in a human-readable format. This approach should consider whether the journal provides any tool for exporting citation metadata, or not, and also, for those providing such functionality, identify the ways in which metadata are provided and establish percentual rates of journals providing each specific metadata format, be it machine-readable metadata, be it human-readable format.

7.4 The provision of citing metadata within headers and footers: Some publishers provide basic bibliographic metadata regarding their articles through headers or footnotes. This information can appear on the first page or in all pages of an article, and usually provide basic metadata regarding both the article and the journal issue for supporting the writing of the bibliographic reference of the article itself. These metadata generally comprise (at least) the journal title, the year of publication, the number of the volume and issue, and initial and final article page numbers. This feature should be quantified into a yes or no evaluation regarding its presence within the article.

7.5 The way of presenting journal titles within bibliographic references: According to the reference style adopted, the titles of the cited works within the bibliographic references may be given in full or in its abridged format. The point is that the abridged version of the journal titles may lead the reader to an ambiguous comprehension of the cited work’s title and, consequently, constitute a barrier for accessing the cited work. Considering this, it is aimed to quantify and identify the subject areas which adopt each of the forms of presenting journal titles in bibliographic references, and the sources from where such adopted abbreviations should be based on.

7.6 The total number of authors: this is a basic data, that possibly will support a discussion on the relationship between the number of authors of an article and the accuracy of the

information given in mentions, quotations, and bibliographic references. the analysis should quantify the total number per article, in order to support the further indication of the average indexes per subject area and subject category.

7.7 The non-textual sources provision: Graphics, photos, and figures also may be included in a text as a mention or as a quotation. The premise here is that not always the source from where these excerpts are extracted or adapted are indicated. The correct procedure is to always inform the source from where external data were taken however, it is not a very common practice among authors. This evaluation should indicate the percentual portion of the articles per subject area which indicates the source for non-textual cited content and the portion which does not.

7.8 The in text-reference pointers format: the final report of the research should include a discussion on the correspondence between in-text reference pointers, mentions and/or quotations, and respective bibliographic references. Because of that, it will be necessary to consider the way in-text reference pointers are presented within the article texts in order to verify: first, whether they are clear and understandable; second, whether they allow the identification of the bibliographic reference that corresponds to the mentioned or quoted passage regarded to the in-text reference pointer; third, whether the data provided within in-text reference pointers correspond to the cited FRBR Work, FRBR expression, FRBR manifestation or FRBR item; fourth: whether the correspondence between in-text reference pointer and its respective bibliographic reference is harmonious in order to make it possible to identify the exact cited excerpt within the cited work without the need to proceed additional and complementary queries in summaries, indexes, catalogs and so on. Whenever any inconsistency between in-text reference pointers and bibliographic references entries are detected, it should be considered an unconformity and, therefore, a barrier to the identification of the cited work within the bibliographic references list. That is to say that in-text reference pointers that do not match any bibliographic reference entry in the bibliographic reference list, even in cases in which this connection may be inferred, should be considered an unconformity, as a mention or quotation not included in the bibliographic references.

7.9 The links between in-text reference pointers and bibliographic references: considered as a facilitating tool for the reader, it should be analyzed whether the in-text reference pointer is hyperlinked to its respective bibliographic reference. Besides being a convenient tool for the reader, it favors the referencing task and explicit the link between mentions, quotations and its respective bibliographic references. Another point still on the same aspect to be evaluated, in cases that in-text reference pointers are hyperlinked with the bibliographic references, is whether the linkage between them is round, that means, by clicking on the in-text reference pointer the reader is submitted to the correspondent bibliographic reference and, by clicking on some point at the bibliographic reference the reader is sent back to the

excerpt of the text where the work, which is represented by the bibliographic reference, is mentioned or quoted.

7.10 The total number of mentions and quotations per article: the total number of mentions and quotations used in each article should be quantified. This data will support a discussion on how the subject areas tend to cite external publications and, which is the most common way of doing it: in a literal form, i.e., as a quotation, or in an interpretative form, i.e., as a mention. No matter the length of the quote, the author's intention in quoting should prevail. Quotations shall be considered as so, regardless of the length of the quoted excerpt, be it a phrase, a paragraph, or an expression, as long as it is properly identified as such, according to the reference style adopted. this analysis is not intended to check quoted passages within cited works so, it should be assumed that the information provided by the authors, i.e., the markup for mentions and quotations, are trustful indications of the inclusion of the content from a cited work. This analysis should only consider textual content. So, although figures, tables, and other non-textual content also may be mentioned and quoted, they should not be considered by this analysis.

7.11 The mentions indicating pagination data concerning the cited work: a mention in characterized by the reproduction of a cited work under the words and the interpretation of a citing author. By using this way of citing, the indication of the page numbers of the excerpt mentioned in the cited work is not mandatory. Since the scientific normalization requires a minimum level of uniformity, it is assumed that the indication of the page numbers for mentions is a choice of the author. However, it is also understood that the indication of the page numbers in one mention requires the indication of this metadata in all mentions within the text, in order to keep the uniformity within a text. This factor can be used to demonstrate the importance and the rigor in which normalization matters are carried out by the publisher. Facing this, there should be quantified the number of mentions per article that indicate data concerning the pages where the mentioned content can be found in the cited work.

7.12 The quotations not indicating initial and final pages on the cited work: quotations are the transcription of a passage from a cited work. As being a literal reproduction of a cited work, the quoted passage should be evidenced within the citing work's text, i.e. marked up between quotation marks or in an indented paragraph. Most (if not all) of the reference styles recommend the adoption of quotation marks (double or single) as a mark up for quotations which should be followed by an in-text reference pointer. For articles adopting the author-date citing system, the in-text reference pointer should provide at a minimum the cited author surname, the year of publication of the cited work and the initial and final page numbers where the excerpt quoted may be found within the cited work. For articles adopting the citation-sequence system, the in-text reference pointers concerning quotations generally provide a number corresponding to the respective bibliographic reference it refers to in the bibliographic references list, followed by the initial and final pages from where the

quoted excerpt was extracted in the cited work. These metadata format presentations are usually mandatory within reference styles' guidelines and configure quite crucial metadata for retrieving the quoted content in the cited works. Besides, the indication of page numbers is one of the factors that usually differ in-text reference pointers for mentions and for quotations. Facing all of this, the in-text reference pointers for quotations should be analyzed within the selected journals in order to identify:
a) the rate or articles per subject area whether the initial and final page numbers where the quoted excerpt may be found in the cited work are given within the in-text reference pointers;
b) the rate or articles per subject area whether it is possible to identify the exact starting and finishing point of the quoted excerpt within the citing work, that is, whether the quotation is presented between quotation marks or indented.
It should be reminded that this method does not foresee any consult to the works cited by the analyzed articles. So, quotations not properly marked up, i.e., those not presented between single or double quotation marks or indented, should be considered mentions providing pagination data, regardless respective in-text reference pointers provide the pagination of the cited passage within the cited work.

7.13 The way quotations are presented in the article texts: The form of presentation of quotations should also be a point of observation, starting from the premise that reference styles do not give enough elements regarding the guidelines for presenting quotations within the articles. Considering that there is no established standard on the length that defines long quotations, which are generally presented with an indent, and short quotations, which are generally presented between quotation marks, it is expected to confirm that there is no uniformity on the way quotations are presented within journals, regarding the use of quotation marks and also indent formatting. In this line, the analysis should quantify the rate of quotations per article adopting each presentation form.

7.14 The average number of mentions for each cited work in the same article: each mention and citation for each cited work should be quantified in order to have an average number of mentions and quotations per article by author, by type of publication, and by subject area.

7.15 The FRBR relation between quotations and bibliographic references: This particular aspect of the analysis starts from the understanding that, in a comparison with mentions, quotations are more evident and precise within texts. Once the pages numbers where the quoted passages can be found in the cited work are usually given within in-text reference pointers, it is attributed a FRBRized feature to them. That means that by giving the page numbers, the in-text reference pointer usually refers to a FRBR manifestation, while the omission of this metadata generally links the in-text reference pointer to a FRBR Manifestation. Since the cited content stands out in relation to the support in which it was published, it does not matter whether the citing author consulted the electronic or the printed version of a work. What really matters in this respect is that the work remains the

same, regardless of the embodiment it may take. Considering that approach, this point of the analysis should identify the relationship established between quotations, its respective in-text reference pointers, and the respective bibliographic references. The initial premises suggests that bibliographic references generally describe the cited works considering the FRBR manifestation level, while in-text reference pointers, not always share the same FRBRized perspective. Facing this, the number of in-text reference pointers referring to quotations corresponding to the FRBR Works, FRBR expressions, and FRBR manifestations concepts should be quantified. In a complementary way, bibliographic references should be considered under the same perspective. then, the results of both observations should be put together to determine the rate of distinction between the level of description observed within in-text reference pointers referring to quotations and their respective bibliographic references, from the FRBR perspective.

7.16 The total number of bibliographic references: Theoretically, the total number of bibliographic references observed in the bibliographic references list should be equal to the number of works cited within the text it belongs to. starting from this premise, the analysis should quantify the total of bibliographic references per article.

7.17 The bibliographic reference lists assortment: Bibliographic references lists should be evaluated while the assortment of bibliographic references and classified into numerically or alphabetically assorted. then the analysis should consider whether these assortments are correct, i.e. the proper alphabetical ordination of bibliographic references for articles adopting author-date system and the proper ascending numerical ordination of bibliographic references for articles adopting citation-sequence system.

7.18 The use of et al.: et al. is the Latin abbreviature of et alii, and is used in bibliographic references and in in-text reference pointers to omit the name of multiple authors, usually when they are more than 3, according to the reference style adopted. Since in-text reference pointers are the main elements denoting citing works and bibliographic references within the text-bodies, the method proposes the observation of the correspondence established between in-text reference pointers and bibliographic references, and the proper use of et al., considering the recommendations of the reference style adopted by the publisher of the journal in which the analyzed article is published. In this perspective, the quantitative analysis should focus on the following points:
a) the number of in-text reference pointers referring to mentions and quotations per article properly and improperly indicating et al.,
b) The number of bibliographic references properly and improperly indicating et al.

7.19 The mentions and quotations without a respective bibliographic reference: all mentions and quotations should have a corresponding bibliographic reference in the bibliographic references list, to allow the reader to identify the cited work. The reverse situation, likewise, is also valid: there should not be a bibliographic reference in the bibliographic reference list without a correspondent mention or quotation within the article text. This aspect should be

considered by the analysis by checking whether all in-text reference pointers considered in the text-bodies clearly correspond to a bibliographic reference in the bibliographic reference list and vice-versa. Such correspondence should be clear, and so, any inference can be admitted in this scenery. So, for all cases in which in-text reference pointers' main access points do not coincide with those considered by the apparent corresponding bibliographic references, should be considered as unconformities which suggest flaws in the normalization process.

7.20 The type of publications cited: the number of bibliographic references per article representing each type of publication (e.g. articles, books, proceedings, etc.) should also be considered. This parameter will support an evaluation of the most relevant type of publications in each subject area. Books published simultaneously in traditional press format and in electronic support, should be counted as printed copies, unless there is an explicit note or metadata in the bibliographic reference indicating the consult to the electronic version of the publication. The bibliographic references that do not give enough elements for the identification of the type of publication to which they refer, should be classified as undefined.

7.21 The modality of access of the cited works: a quantitative evaluation on the modality of access of the cited works should support a discussion on the relationship between consulting an open access publication and including a note regarding this in the respective bibliographic reference. This analysis aims to verify the veracity (or not) of the premise that authors generally cite open access publications and do not include any note about it in their bibliographic references. Each bibliographic reference included in the bibliographic references list should be evaluated and classified into free access or restricted access. Hardcopy sources should be considered as restricted access since the concept of freely available refers to the online and unrestricted access to the publication, free of charge. Books and other types of publications released both in printed and electronic versions should be considered as restricted access unless there is an explicit note in the bibliographic reference regarding the consult of the electronic version by the author.

7.22 The DOIs and hypertext links: Since DOI metadata, e.g. DOI numbers and DOI Hyperlinks, refers to the unique identifier for an electronic publication available online, it is convenient to indicate it in the bibliographic references whenever it is possible. The same concept applies to publications that are available online, under a hypertext link, although they are subject to change, contrary to DOI data. However, there is a premise that DOIs metadata and hypertext links are rarely included in bibliographic references, even developing such a crucial role in information retrieval. In order to verify the validity of this premise, a quantitative analysis of the indication of hypertext links and DOI metadata within the bibliographic references will be carried out. It should be quantified the following aspects:
a) the number of bibliographic references not including hypertext links or DOI metadata, considering only bibliographic references referring to freely available online works,

b) the total number of bibliographic references including hypertext links, considering the whole bibliographic references.

7.23 The link rot issues: Electronic resources available on the Internet are subject to be reallocated or made unavailable permanently. Consequently, the links originally pointing to these resources lost their functionality, and do not permit the access to the file, web page, server, or another source to where these links were originally pointed at. This phenomenon, called link rot, is frequent among bibliographic references and this method intends to quantify the incidence of these link rots within the bibliographic references lists of the articles composing the sample. So, for each hypertext link or DOI hyperlink included in the bibliographic references should be tested by clicking on them and observing the webpage to which such link is pointed to. Then, this analysis should evidence:
a) the frequency in which bibliographic references include a hypertext link rot,
b) the frequency in which bibliographic references including DOI hyperlink rots and,
c) The number of hypertext links or DOI hyperlinks pointing to webpages other than the one originally containing the cited work, i.e. library catalogs and databases.

7.24 The format of bibliographic references numbers at the bibliographic references list: for the articles adopting the citation-sequence system, the format of the number designating each bibliographic reference in the bibliographic reference list should be observed. These data should provide an overview of the range of formats adopted by publishers to express the same data within bibliographic references.

7.25 The bibliographic references metadata compilation: each bibliographic reference must be analyzed individually under the aspect of the metadata set provided. Thus, the analysis must indicate which are the descriptive elements usually considered by each journal on the bibliographic references for each type of publication. The method should consider and classify each descriptive element individually, such as "included in the bibliographic reference" or "not included in the bibliographic reference", in order to obtain a list of the description elements indicated in the bibliographic references by journal, by article, and by subject area. This list should be the starting point for delimiting the elements of description that are indispensable for the identification of a publication by means of a bibliographic reference which from now on will be called “essential metadata set". This analysis should provide the percentual of articles per subject area considering each descriptive element within their bibliographic references. descriptive elements considered by at least one bibliographic reference within an article should be classified as "included in the bibliographic references".