

International Journal of Human Sciences Research

REFLECTIONS ON THE IMPACT OF THE WEB ON HISTORY OF SCIENCE WORKS: “FAKE DATA”, TRANSPARENCY AND SUBJECTIVITY

José Adolfo S. de Campos

``Observatório do Valongo``,
``Universidade Federal do Rio de Janeiro``
<https://orcid.org/0000-0003-4543-2280>

Jorge Fernando Silva de Araujo

``Colégio Pedro II``, Rio de Janeiro
<https://orcid.org/0000-0002-6702-4184>

All content in this magazine is licensed under a Creative Commons Attribution License. Attribution-Non-Commercial-Non-Derivatives 4.0 International (CC BY-NC-ND 4.0).



Abstract: The creation of the WEB, one of the most popular services on the Internet, generated a revolution in communications as impactful as Gutenberg's movable type machine. Technological advances over the last 50 years have made global communications almost face-to-face. The presence of the WEB was felt in all areas of human knowledge, especially in Education and Science, offering support for the development of new projects. The objective of this text is to reflect on how the characteristics of the WEB impact the performance of intellectual work, in this case the creation of an article on a topic in the History of Science. Based on the authors' personal experience in the area, the stages of the process of preparing an article were examined. It was found that searching for references via the WEB can flood the historian of science with references that are not relevant to the objective of the work – we call it “Excess”. Suggestions are presented to make the process of separating “Excess” from significant references more effective. We also address the influence of “Fake Data”, “Transparency” and “Narrative” in choosing the characteristics of the type of audience to whom the work is aimed. We concluded that a good choice of references is fundamental, but the big problem is that there is no criterion for valuing references.

Keywords: History of Science, WEB, Fake Data, Fake News, Transparency.

INTRODUCTION

A historian of science studies the evolution of an area of science, such as mathematics, astronomy, physics, etc. The nature of the studies and research carried out leads them to seek information in books, periodicals, newspapers, handwritten documents, photographs, music, films and videos, maps, paintings, buildings, etc. These documentations are usually located in institutions that are in different cities or even countries. Furthermore, they can be written in several languages. Until the advent of the WEB, the non-availability of time and/or financial resources (which are normally scarce), meant reducing the search for documentation to a few places or changing the focus of the research, adapting it to the available resources and the result.

The research is communicated through books and articles published in specialized magazines and periodicals.

The WEB, which emerged in 1992 at the proposal of British scientist Tim-Berners Lee, makes use of the Internet¹ which is a worldwide network of interconnected computers and through it, data and information are transmitted to any user who is connected. THE WEB² is an Internet service that interconnects users from anywhere in the world, allowing the transfer of text, sound, image and video files instantly.

The connections between practically every place in the world and the almost instantaneous propagation of information led to the creation of a gigantic and growing **Global Digital Database (BDDG)**, accessed through query engines such as Google (www.google.com) and Bing (www.bing.com). Consulting this Digital Database can provide answers in fractions of a second about topics

1.; In September 2022, the internet had more than 5.4 billion users and connected 2 billion websites (<https://www.internetlivestats.com/watch/>).

2. Analogy used for the relationship between WEB and Internet: The internet is the roads that connect cities, while the WEB is the buses with data that travel on the roads taking and bringing information.

or events you want information about, which are stored on websites. From this BDDG you can also create another database with specific purposes, such as a **Bibliographic Database (BDB)**, made up of book collections that are stored in thousands of libraries around the world. Although connected to the network, most libraries still do not make book content available for remote consultation. The Google Books search engine (<https://books.google.com.br/>) searches for full texts of books available on the internet. There are also websites that allow access to documents from public archives and archives from museum libraries, universities and cultural institutions.

The worldwide reach of the WEB and the ease of handling multimedia files have made it a great tool for historical research, however, care must be taken when analyzing the results of searches in Databases. The objective of this article is to discuss the impact of using WEB resources on the creation of History of Science works and additionally to draw the attention of science historians to the facilities introduced by the appearance of the WEB.

WEB FEATURES

There are several characteristics of the WEB that help historians of science when choosing the object of their research and during the various stages of development of the work. The almost instantaneous speeds of dissemination and propagation of information have made the possibility of “in-person” meetings using all multimedia resources real, whether for discussions in “petit committee” or for presenting work at Congresses. The dissemination of information via the WEB covers the whole world because it uses the Internet network, which is installed in every corner of the Globe. The propagation of information is also practically instantaneous. These features plus the possibility of sending

3. Some publishers adopt the Creative Commons licensing system (<https://creativecommons.org/>).

4. See (BURKHARDT, 2022).

text files with multimedia (sound, image, video) allow real-time communication between researchers, exchanging information and files, encouraging collaborative work.

The resources available for editing texts, such as the adoption of “templates” where authors only have to follow the instructions so that the articles comply with the magazine’s standards, have had a great impact on the editorial environments of magazines and periodicals scientific, speeding up the publication process, greatly reducing the delay between text submission and publication, a constant complaint among researchers. Due to the low cost of producing texts, some of the printed magazines launched electronic versions and magazines with only the electronic version appeared, published according to the “Open Access” protocol.³, in which access to content is free. A large number of electronic journals do not charge for publishing articles, but there are still journals that charge a subscription, a practice that makes it impossible for scientists from poor and developing countries to publish articles.

The use of the WEB has accelerated the stages of documentary and bibliographical research, but has not yet replaced the “on-site” search for primary documentation in archives and libraries, because there is a lot of documentation outside the network, without scanning or copying available.

The lack of control over the content of what is posted is one of the characteristics of the WEB, which arouses controversial feelings, as there are arguments for and against control. Critics argue that the total lack of content control on websites facilitates the dissemination and spread of false news (Fake News), which can harm the reputation of people and companies by leading to wrong conclusions.⁴. Those who support the measure argue that it avoids the political censorship

exercised by authoritarian and oppressive governments, allowing minority groups to express themselves and denounce arbitrary acts. The name “Fake News” is a version boosted by the “instant” dissemination and propagation characteristics of the WEB, of “rumors” and “rumors” as they were formerly called. An “impressive” “farce” was the theatricalization of the War of the Worlds, a novel by HG Wells, made by Orson Welles in 1938, which was broadcast on dozens of radio stations in the United States, giving the impression that the invasion of Earth by Martians had begun leaving millions of people terrified by the “fake invasion”. Sometimes things get out of control. As texts can be published directly on the WEB without anyone’s intervention, the user can write any nonsense, such as stating that “the Earth is at the center of the Solar System”. The WEB is home to a “gigantic” amount of texts with errors and inappropriate additions, especially harmful when it comes to science and science communication topics. However, the freedom to write ideas and disseminate them has many more positives than negatives and, in practice, it is very difficult, if not impossible, to define a line that separates “censorship” from what society considers “responsible freedom.”

Another point is the issue of resilience to incorrect or false information on websites. As the Internet was created as a network structure, the interruption of one route does not prevent information from reaching its destination via another route. Therefore, removing a website from the network is not an easy operation, neither from a technical nor a social point of view. Sites whose owners have “forgotten” or “died” remain on the WEB “ad infinitum”, unless the person responsible takes them offline. There is no cleaning company to eliminate the millions of inactive websites⁵.

5. An estimate indicates that the number of inactive websites is around 83% of the total (<https://programadoresdepre.com.br/quantos-sites-existem-em-todo-o-mundo/>).

6. To have a broader and more detailed notion of what the work of a Science historian is, see Martins (2005).

Added to this, as it is very easy to copy a website, the information we want to delete/correct often ends up being reproduced on other websites, without identifying the origin. This multiplicity of copies of website content, in whole or in part, creates difficulties in validating statements based solely on quantity.

INFORMATION SOURCES AND THE WEB

In general, the historian of science, when proposing to carry out research on past facts or events, has to complete several steps to reach his objective, which may be a thesis, a book or an article that will be published in Congress annals. or scientific magazines in the area, reporting their discoveries and/or reflections on the chosen topic⁶. The activities can be briefly described in 7 steps (Chart 1). The last column shows whether the WEB is used. The codes are: T – Maybe depending on the case; S – Yes; P – Partially.

Once the research object has been chosen (1st stage), which may or may not use the WEB to assist, he must think about the existence of potential sources of historical information (2nd stage), where he will find documentation related to the topic and even more so, if these Documentation sources are available to be searched. For this task, he can have the help of the WEB, which will indicate the website addresses of institutions with documentation of interest to the research topic. Unfortunately, some collections of institutions that could be precious sources of documentation are not available for various reasons such as: unorganized files piled up in rooms, files containing extremely degraded documents and files with unauthorized access due to “excessive zeal” in handling the documentation. The state described refers to some Brazilian institutions, but it can

certainly be applied to institutions in some other countries.

Phases	Developed activities	WEB
1st	Choosing the object of historical research	T
2nd	Carry out a preliminary survey to check whether there are accessible sources of information about the chosen research object	s
3rd	Once the research object has been defined, start searching for documents and/or bibliographies in WEB databases	s
4th	Carry out "on-site" searches for documentation in local archives of information sources	P
	<i>The searches defined in the 3rd and 4th stages interact with each other, feeding each other back.</i>	
5th	Classify information sources by relevance	s
6th	Write the text of the work following some precautions	s
7th	Send the text within the standard requested to the magazine	s

Table 1: Activities developed by a Science Historian

Another type of difficulty is obtaining copies of documents, either through charging fees or through bureaucracy to obtain permission to copy. In some places you can get pencils and paper to copy. As documentation will inevitably be destroyed by the action of time, it is necessary to invest in the digital reproduction of documents, which is also not eternal.⁷

In the last two decades, mainly, there has been an increasing increase in the digitization of historical files and their consequent availability through the WEB. This movement allowed researchers access to many primary source collections⁸, located in various parts of the world, of which one would not otherwise be aware or which would be out of reach due to logistical and economic difficulties – distant locations, lack of funding for travel and long periods of research outside their institution. In Brazil, we have a program that

7. See problems with long-term file maintenance (ARELLANO, 2004).

8. They present the information in its original form, without interpretation or analysis by other authors (Martins, 2005, p310).

9. The SAO/NASA Astrophysical Data System (<https://ui.adsabs.harvard.edu/>) has more than 15,000,000 records of Astronomy and Astrophysics publications; the SIMBAD base (<https://simbad.unistra.fr/simbad/>) has information on about 13,500,000 astronomical objects - 2021 data.

deserves mention and applause, the Program of Brazilian Digital Newspaper Archive (<http://memoria.bn.br/hdb/periodico.aspx>), from the National Library, which has already digitized and accessible via the WEB, around 50 million newspaper pages from 1800 onwards, which constitutes a precious repository of information.

Although document research via the WEB has greatly facilitated the search for copies of original documents, the "on-site" search for documents in institutions is still the most important, time-consuming and tiring phase. Gathering the bibliography on the topic to be studied was greatly speeded up by the search mechanisms provided by the WEB, making the retrieval of relevant information faster and less tedious. The offer of information sources has certainly been greatly expanded with the use of resources available on the WEB, but on the other hand, this offer presents many copies and repetitions of articles, sometimes with errors and inappropriate insertions, which can lead the researcher to erroneous conclusions.

The area of Astronomy stands out positively in this regard as it created, under the command of the International Astronomical Union (IAU), gigantic databases bringing together copies of publications and astronomical observations since the 19th century.⁹, greatly facilitating the search for detailed information on astronomical objects (stars, galaxies, interstellar medium, nebulae, etc.).

The historian of science, when analyzing the documentation collected from documentary and bibliographical research, seeks to separate the documents and publications that can effectively contribute to the development of the chosen theme (3rd and 4th Stages). This

mass of information presents, in addition to pertinent data, data with omission of the origin of the information, duplicate information from the same website, duplicate data from “mirror” websites, factually contradictory information and “false” data (*Fake Data*). We define “*Fake Data*” as “data that originate from documents where they occur, with no intention of falsifying them on the part of the researcher, transcription errors, translation errors, printing errors (typing of dates, numbers, names) and the repetition of erroneous data other documentation. Unlike the “*Fake News*”, which would be news purposely inserted on the WEB with the aim of confusing and leading the reader to erroneous conclusions, based on manipulated data, the “*Fake Data*” occur without the explicit intention of inducing errors.

Searches almost always present an excess of sites with irrelevant information, because depending on the form of the question to the database, sites containing documents and information that do not fit the purpose of the search are selected. Let’s call it excess of references to websites that must be removed from the sample, normally consisting of homonyms, names of places and streets, partial names of companies and people, events outside the research period and websites with marginal information irrelevant to the desired objectives. A careful choice of search parameters¹⁰ can eliminate a large part of the “Excess” from the sample.

As an example, searching for information on Google about the Bahian scientist “Manoel Pereira Reis” written in quotation marks¹¹ (returned 17,500 references), leads to different results than when written without quotes (returned 9,100,000 references).

10. See “Refine WEB searches” (<https://support.google.com/websearch/answer/2466433?hl=pt>) and “30 tips and tricks to improve your Google searches” (<https://canaltech.com.br/internet/dicas-e-truques-para-melhorar-as-suas-buscas-no-google/>).

11. Only select sites that have the full name as it is written in quotation marks.

12. Hence the importance of having a low number of websites to examine in detail.

13. This simple division was adopted, but the taxonomy of historical sources is a subject under discussion (BARROS, 2012).

Examining the resulting sample, it is seen that there are many citations to **Road** Manoel Pereira Reis. Let’s remove these sites that have Rua in the name of the reference (“Manoel Pereira Reis”-**Road**), which now results in 6950 references. Pereira Reis is known for helping to design the constellations on the Brazilian flag. As you are not interested in this topic it will be removed (“Manoel Pereira Reis” -Rua -**Flag**) resulting in 2880 references. The well-guided search process reduced the number of references from 17,500 to 2,880, which is still a high number. Ideally, the number of references examined must not be high (the authors’ personal experience recommends that it must not exceed 100). At this stage, the filter must be applied via the search engine.

WEB PROBLEMS: FAKE DATA, TRANSPARENCY AND SUBJECTIVITY

At the end of the previous stage, the set of references still had “Relevant Data”, “Fake Data” and Excess. In this stage (5th stage) you will use the **individual inspection**¹² of references to eliminate websites with duplicate references (website with the same domain) and with partially or completely duplicated content (“mirror” websites), identified by pieces of text with the same phrasing, and classify the sample references by relevance. The classification of references by relevance can be on two levels. At the first level, a screening is carried out to classify the sources of information¹³ in primary, secondary or tertiary sources. Sources depend on their originality and their proximity to the source of origin (called Primary). The Primary Source is the source that was created at the time in which it is studied, by a source of

authority. The Secondary Source is the result of discussion of primary source material.

The Tertiary Source¹⁴ is a compilation of primary sources and secondary sources. At the second level, a more careful analysis of the primary sources is carried out to give an order of reliability, which in reality is an order of preference.

The reduction of “Fake Data” is another step in the process to increase the reliability of the data used in the construction of texts. However, “Fake Data” is more difficult to discard because the errors are accidental and random and the historian of science needs to read and check the data within its reference. Imagine the phrase seen in the original Ministry of Agriculture report: “The Astronomical Commission of the Ministry of Agriculture, Commerce and Public Works was abolished on January 4, 1879”, but due to a copying error the date “January 3, 1879” was written. But how to minimize “Fake Data”? Historiography indicates that works must rely as much as possible on documentation obtained from primary sources¹⁵, whether collected by the researcher directly from the original source or from citations from secondary sources. When obtained from secondary sources, the historian of science must try to access the cited documentation to verify the fidelity of the factual information.

It is not uncommon for documents originating from primary sources to differ on factual issues. In this case, it is up to the researcher to decide which of the sources, based on an analysis anchored in arguments explicitly stated in the work. A *loyalty* factual evidence supported by primary source documentation is important to reduce or eliminate “*Fake Data*” of a research/work.

For example, for the article about Manoel Pereira Reis it is necessary to describe his life

from his birth in Bahia (1837) until his coming of age in Rio de Janeiro (1858). In the search for references we found four documents (secondary sources) that mention that Pereira Reis was born in the city of Salvador and two from supposedly primary sources: one quotes of S. Sebastião (from Passé?) by professor at the Polytechnic School and colleague Francisco Ferreira Braga and another by friend and former student Américo Brasília Silvado who indicates Cachoeira. A preliminary analysis ruled out Ferreira Braga as a source of authority because there was no evidence of greater proximity to Pereira Reis. Exactly the opposite occurred with Brasília Silvado as he was a very close friend of Pereira Reis and his family. Using this example, you can see how non-trivial it is to classify the relevance of a source and how much the classification depends on subjective factors that influence the historian of science.

Another fundamental point in the process of confirming the fidelity of the data cited in the work is the *transparency* in the referencing of information, that is, the correct citation (indicating the full path) of the documentary references that anchor the facts and conclusions of the work (DANUELLO et al, 2023). Sometimes the complete path to get to the cited document is omitted, making it difficult or even preventing another researcher from checking the information.

Some works intended for the Academy’s public, intentionally or not, fail to clearly indicate where the documentation is located¹⁶. Among the reasons for the author not indicating the complete path to locate the original document (or copy thereof) are: not attaching importance to the citation, because he thinks no one will need this documentation; think that an incomplete citation is enough for other authors to locate the documentation;

14. Examples of tertiary sources: Directories, Portals, Bibliography, Encyclopedias, Catalogs, Wiki Sites.

15. (https://pt.wikipedia.org/wiki/Fonte_primary) accessed on 2-09-2023.

16. It was common for works, until the beginning of the 20th century, not to present a bibliography with the path.

no longer knows how to describe in detail where the document is located due to the lack of detailed annotation¹⁷; do not want to share their discovered “mine map” in order to use the data exclusively (this is an ethical problem). It's also an **ethical problem**, intentionally extracting sentences from documents, ignoring the context in which they were uttered and distorting the meaning in order to justify the argument.

While the *loyalty* to the facts and *transparency* in referencing sources can be directly compared, a third component - the analysis of events and the consequent elaboration of a **narrative**, is totally dependent on the author and his views on the chain of situations addressed and which may affect the choice of primary sources. A *subjectivity* in the analysis depends on the moral values, beliefs and life experience of the author. As a consequence, the same data set can give rise to different narratives by different authors.

As for writing the text (6th stage), some care must be taken by the historian of science, who ideally must be impartial and maintain a certain distance from the research object. The idea that scientists/researchers would have a position of neutrality, remaining exempt, without external influences, is a myth (MARTINS, 2001, p. 40). In principle, the further away from the facts they are spatially and temporally, the less influence researchers would suffer, even so, the system of beliefs, prejudices and individual values would influence the analysis of events (MARTINS, 2001, p.38).

Another common mistake is trying to judge past events using concepts and values from the present (Anachronism) or from a particular culture. This introduces a bias into the narrative, which is detrimental to the result obtained (MARTINS, 2001, p.38).

17. With the progressive introduction of standardization for archives (NOBRADE – Brazilian Standard for Archival Description, 2006), the locations of some documents had their identifiers changed, making their access difficult.

18. See more recommendations on work methods and ethics in Martins (2001, p38-39).

The “uncertainty” that permeates the narrative cannot be “reduced” because it is a subjective issue, that is, it will vary from author to author, and there is no “truer” interpretation of the past or an “absolute truth” (MARTINS, 2001, p.38).¹⁸

So far, only suggestions have been presented for the execution of a script intended for the production of academic texts. However, taking advantage of the research effort, the Science Historian can also write texts aimed at different “audiences”.

TEXTS FOR ENTERTAINMENT, PUBLICITY AND ACADEMICS

When producing texts on historical themes or events, the author must clearly keep in mind the intended target audience, since there are varying levels of requirements regarding fidelity to the historical facts narrated, transparency in citations from sources of information and adequate narrative. In general, the “audience” can be classified into three categories: general public with varying levels of education and culture; public with specific interests in historical and/or scientific themes; and academic audiences with specific interests.

The text of a work is made up of objective elements and subjective elements. Objective elements (such as names, dates, numbers, *ipsis litteris* copies of other texts) are subject to verification as to their fidelity to the original documents; subjective elements (such as the content of comments and opinions) are not, because they depend on the author's value system.

The texts used in the production of literature intended for the entertainment of the general public (historical novels, soap operas, films, theatrical texts, etc.) have no commitment to fidelity to the facts or transparency of

information, and the narrative aims to captivate the public introducing fictional conversations and facts.

For the public with a specific interest in historical/scientific themes, the objective is to present the theme for science dissemination (books, documentaries, texts on the Web, texts in newspapers and periodicals, public lectures, etc.), where the narrative has to be attractive and informative, but faithful to historical facts. To make the text more fluent, details and facts that are not relevant to the proposed objective are generally omitted, and transparency is also sacrificed for the benefit of fluency. Fidelity to the facts is the positive point. However, the narrative tends to be very superficial due to the space/time available in the chosen media.

One of the main problems of popular science texts is the author's scientific credibility. Scientific credibility is sometimes confused with the author's celebrity in the media. In other words, popularity is not a factor to be taken into consideration, when it comes to scientific statements, that is, a high number of "posts" or followers on a social network does not mean credibility.

Texts intended for academic audiences (books, dissertations, theses, articles published in specialized journals, etc.) need to be faithful to the facts, have transparency regarding the complete referencing of sources, and in addition to have a narrative supported by the facts presented. Table 2 presents a summary of the characteristics that each type of text must have depending on the target audience.

CONCLUSIONS

The emergence of the WEB/Internet had a major impact on all areas of human activity, revolutionizing communications and our way of life. Computers, cell phones and the internet are part of our daily work and leisure lives. We became "Net Dependents". For historians of science, the new services provided by the WEB provided the opportunity to: increase the scope of historical research work; greater depth in searches for primary sources; greater interaction between researchers; greater access to academic articles and easier publishing of your work. As a counterpoint, there is a need to establish criteria for valuing information from websites and eliminating unreliable sources, with erroneous data and/or duplicate data. Table 3 briefly presents the impact of the emergence of the WEB on works in the History of Science, which are discussed in more detail below.

a) Due to the network design of the internet, information is disseminated across the globe (globalization), with an almost instantaneous speed of propagation. The WEB added resources and multimedia, making communication between researchers much easier, making "live" meetings almost commonplace, increasing the chance of establishing collaborations with the exchange of information and possible help in searching for documentation in distant archives. These collaborations increase productivity and save time and money;

b) In academic life, the publication of articles and works, as well as participation in Congresses with the presentation of works, is vital for the researcher's appreciation in both the scientific and economic aspects. Scientific journals were printed on paper

Requirements →	Fidelity to the Facts	Citation Transparency	Narrative
Target Audience ↓			
General public	There is no obligation to use true data.	There is no obligation	Free. Introduction of fictional events and dialogues
Audience with specific interest	Use primary sources of information whenever possible.	There is no obligation	Can be with omission of events and simplified
Academy Audience	Always use primary sources of information.	It is necessary	Rigorous description of the facts

Table 2 – General Characteristics of Texts on History

Characteristics		Consequences
Resource	Qualifier	
Dissemination of information	Instant	Increased chances of collaboration with other researchers
Information propagation	Instant	Increased chances of collaboration with other researchers
Ease of Communication	High	Increase in “virtual” meetings between researchers
Multimedia in Communication	Easy	Increased meeting productivity
Text writing	Easy	Anyone can write poor quality texts with errors
Content control	Difficult	Increased rate of proliferation of websites with misinformation. Difficulty distinguishing common academic websites
Text Publication	Easy	Difficulty distinguishing the relevance of the text
Publishing Costs	Lows	New electronic magazines; texts released free of charge for reading and publication; increasing diversity of origins
Information Resilience	High	Difficulty eliminating information with errors and dated information
Garbage in Information	High rate	Huge number of websites with errors, mirror copies, irrelevant and superficial
Database – Creation	Easy	Creation of researcher database. Acceleration of the information gathering process
Database – Query	Quick	Acceleration of the information gathering process

Table 3: Characteristics of the WEB x Consequences

and the cost of publishing the article was not cheap¹⁹especially when considering the economic level of poor countries. Publishers said that the high costs of producing and distributing printed material did not allow magazines to be maintained at lower prices. The result is that the researcher often did not have his article published due to a lack of money from his institution.

With the arrival of the WEB, publishing costs were dramatically reduced and material distribution costs practically came to zero. The length of the interval between text submission and publication, which was often a reason for complaints from authors, has been reduced.

“Electronic magazines” emerged, with very low production costs and agility in the production and distribution of scientific articles. Electronic editions provided researchers with free access to articles, without charging fees, using the “Open Access” and “Creative Commons” protocols. With the democratization of information, the production and dissemination of work accelerated, creating new research centers;

c) The nature of the WEB/Internet has allowed the creation of a gigantic and ever-growing virtual database, fed by information produced by billions of users from all parts of the Earth and all cultures. Although the majority of its

19. It must be added that the institution had to subscribe to the magazine, whose annual subscription was expensive.

records do not have content usable by the scientific community, they still constitute a huge mass of data that needs to be explored;

d) The credibility of information sources and their apparently excessive abundance, resulting from searches on the WEB, is one of the major problems faced by historians of science. It is necessary to “separate the wheat from the chaff” and to do this we need to establish criteria that are as comprehensive as possible. The sample needs to go through screening to eliminate non-relevant references, multiple references to the same website, duplicate information from “mirror” websites and seek to minimize errors due to “Fake Data”. It is not possible at the current stage to dispense with in-person visits to archives and libraries that would potentially contain relevant documentation.

e) Bibliographical surveys can also be carried out using the internal search of large libraries, such as the largest in the world - Library of Congress, USA - which has a collection of 155 million copies, of which more than 32 million books²⁰. The overwhelming majority of library collections do not have the content of the books available for consultation;

f) History is one of the areas of the WEB preferred by the general public and where academic research texts, publicity texts, romanticized texts about historical events, humorous texts about historical facts, alternative texts with misinterpretations of facts and “Fake Data” are present. In addition to informal conversations about monuments and historical events, etc. The WEB is a free territory where texts of all types and for all purposes can be

written, without censorship (or nearly so). But, as the saying goes “there are no bonuses without costs”, there is a lack of criteria for valuing texts that really have academic merit and academic texts are confused with “popular” texts which, in most cases, are incomplete reproductions, with erroneous data and absurd interpretations of historical events. There are still no website evaluation criteria. It is up to the researcher to give priority to references whose managers are research institutions, universities, institutes and public archives and preferably which provide access to copies of documents;

g) The resilience of the information contained on websites generated an additional problem when some data, which is widely disseminated, is replaced due to scientific work. Imagine, absurdly, that historical works showed that the Proclamation of the Republic was not on November 15, 1889, but on November 20, 1889. A Google search says that there are around 700,000 websites with the date 15 of November and there would be a single website with the date November 20th. The information will only be corrected by some websites and will remain as November 15th on the network. The updating of historical information is slow (if it occurs) and the historian of science has to be careful with this type of information when screening;

h) When writing a text, the historian of science must consider the Target Audience for whom the text is directed and the parameters Factual fidelity, Transparency in references and adequacy of the Narrative to the facts. Verifying the fidelity of the data provided by the bibliography is very important and historiography recommends that,

20. Top 10 largest libraries in the world (<https://top10mais.org/top-10-maiores-bibliotecas-mundo>).

whenever possible, primary sources of information be used to construct a narrative. If this verification is not possible, the secondary source of said documentation must be clearly cited. A bigger problem occurs if the secondary source uses data from another secondary source (often unknowingly), providing a wider spread of erroneous factual data. The result is a growing number of “sites” conveying the wrong data, which becomes “truth” because it is replicated by everyone.

Transparency in citing possible paths to finding documents, which are sometimes tortuous and difficult to track, is basic for good academic work to be done. In works carried out until the beginning of the 20th century, there was no notion of complete citation. The narrative in academic texts must be faithful to the facts, but must not be boring. Most likely, the historian of science will at some point have to make subjective assumptions, based only on his belief and/or evidence, without the support of factual data.

REFERENCES

ARELLANO, Miguel Angel. “Preservação de documentos digitais”. *Ciência da Informação*, v.33, n. 2, p. 15 – 27, 2004.

BARROS, José D’Assunção. “Fontes Históricas: revisitando alguns aspectos primordiais para a Pesquisa Histórica”. *Mouseion*, n. 12, maio-ago/2012, pp. 129--159.

BURKHARDT, Joanna M. “Chapter 1. History of Fake News”. Disponível em: <https://journals.ala.org/index.php/ltr/article/view/6497/8636>. Acesso em: 12-03-2022

DANUELLO, Jane C., AMADEI, José Roberto P., FERRAZ, Valéria C. T. “ABNT NBR 10520:2023 Guia para elaboração de citações em documentos”. USP, Campus de Bauru, Ago 2023. Disponível em: <https://usp.br/sddarquivos/arquivos/citacoes10520.pdf>. Acesso em: 22-09-2023

MARTINS, Lilian Al-Chucyr Pereira. “História da Ciência: Objetos, Métodos e Problemas”. *Ciência & Educação*, v.11, n. 2, p. 305-317, 2005.

MARTINS, Roberto de Andrade. “História e História da Ciência: Encontros e Desencontros”. *Actas do 1º Congresso Luso-Brasileiro da História da Ciência e da Técnica*, p. 11- 46, 2001. Évora: Universidade de Évora.

In texts for the academic public, the requirements are naturally greater (Table 2), expecting them to have factual fidelity (without Fake Data), which is not always possible to guarantee due to the impossibility of having access to some original documents; transparency in the reference to the origins of documents used as documentary support for the narrative of the historical event.

As it can be seen, Science Historians have a lot to gain if they know how to explore, with knowledge, the services offered by the WEB, especially in the area of gathering information sources. The use of the WEB in searches speeded up and extraordinarily increased the number of references. In some cases, we go from drought to data flooding. The WEB offers the tools, but the historian of science is in charge.

FINANCING

This work was carried out with the support of the Coordination for the Improvement of Higher Education Personnel – Brazil (CAPES) – Financing Code 001.