ABSTRACT

Introduction: The 21st century is marked by the exponential era of data available daily on the web. This era of knowledge is characterized by the pressing need for new technologies to manage the brutal amount of data. In health, it is no different: scientific and technological information needs to be identified, extracted, treated and made available the essential information for decision making. The objective is to contribute to knowledge management in times of Big Data in health with COVID-19. Outline: The work was carried out using a retrospective analysis of the big health data opened for COVID-19. Results: They demonstrate the great worldwide effort in science and technology for the topic in question. The Medline database features over 36,000 articles available. In the European Patent Office base there are 2,871 family of patents, with 3 patents being made available and the pharmaceutical company Pfizer leading the technological research. Implications: It is concluded that the translation of knowledge has been occurring in an urgent way, since, although there is no vaccine or medicine for the cure so far, never in history has there been so much available volume of science and technology in such a small space of time. time, generating scientific and technological production for the advancement of science.

DESCRIPTORS

Public Health; Big Data; Knowledge Management; Coronavirus Infections; Coronavirus.
INTRODUCTION

The 21st century presents itself as an era with 40% of the population connected to the Internet. In this sense, O’Reilly suggested the term Big Data as a gigantic database updated in real time, which easily reaches thousands of terabytes of storage in different formats. Traditional relational database management systems cannot handle these large masses of data. Big Data drives a new generation of methodologies developed to extract economic and strategic value from a large and varied volume of data (structured and unstructured), allowing high speed capture and analysis.

Big Data refers to the third generation of the information age. Initially, this exponential volume of data met the criteria of 3 Vs: Volume, Variety and Speed; 9 later, 2 Vs were added: the attributes of Veracity and Value. Some authors even attribute the last 3 Vs, such as Veracity, Versatility and Viability, where the combination of all "Vs" generates the "V" of Value. Big Data is divided into a perfect data storm, a perfect convergence storm and a perfect computing storm, the latter resulting from 4 phenomena: Moore’s law, mobile computing, social networks and cloud computing. This data collection must be treated to present information searched in a selective and objective way to increase business intelligence, besides allowing an improvement in the decision making process.

As we reflect on health, it is considered as a global public good: that it is not exclusive, that is, that no one or any community is excluded from its possession or consumption; and that its benefits are available to everyone. There is also the apparent consensus that health is not competitive, and that there is no rivalry, that is, the health of a person cannot be at the expense of excluding other people.

In the health spectrum, it carries challenges and opportunities in the globalization process, which is the catalyst for the evolution of the term “Global Health”. Global health can be understood at the same time as a condition, an activity, a profession, a philosophy, a discipline or a movement. However, it must be considered that there is no consensus on what Global Health is, nor a single definition, and its field of action has imprecise limits; however it is indisputable that we are living Health in times of globalization.

Thus, it is necessary to seek to identify, extract and treat the Big Data of Health in this globalized world, in order to focus on the essential information for the decision makers of the present century. Nevertheless, the management of this knowledge is not considered trivial, since the Open Science approach is imminent. Open science is a model of scientific practice that, in line with the development of digital culture, aims to make information available in a network opposite to closed laboratory research. The term also refers to the generation of research materials shared openly, without the need for patents. In this context, the European Community has shown maturity in this area with the promotion of Open Science, as well as some institutions in Brazil, such as, for example, Fiocruz’s proposal for Open Science in its Terms of Reference: management and opening of data for research at Fiocruz.

In this way, organizations that deal with health research have to seek better management of the knowledge management of Big Data in Health, increasingly in an open science, to portray a collaborative and constructive intelligence for society. Therefore, this work aims to contribute to knowledge management in times of Big Data in health and with COVID-19; through an overview of scientific and technological information.

The methodology used was to identify the state of the art in indexed databases such as PubMed, Scielo, Scopus and Web of Science. Likewise, official websites such as the World Health Organization (WHO) and the Brazilian Government’s Transparency Portal (https://transparencia.registrocivil.org.br/registral-covid). The keywords used in the COVID-19 spectrum were “Coronavir * or MERS-CoV * or SARS-CoV * or COVID * or HCoV or SARS or MERS”.

Rev Pre Infec e Saúde. 2020;6:10659
revistas.ufpi.br
INFORMATIONAL EVOLUTION OF DATA AND COLLABORATIVE INTELLIGENCE FOR GLOBAL HEALTH

It can be considered that the development of an investigation is no longer linear. The speed of this evolution has a fast pace. The pace of change of paradigm, of new ideas has been very fast in the world of the knowledge era. It is noteworthy that the first changes, although apparently rapid, took years to develop. It can be exemplified by the fact that the sequencing of HIV (human immunodeficiency virus) took 15 years, while the sequencing of SARS (severe acute respiratory syndrome) was obtained in just 21 days. Nevertheless, the same occurs in COVID-19; where it can be seen that in just about 6 (six) months, the sharing of networked data and scientific and technological information grew exponentially, as will be discussed in the next item.

With technological advances in all areas of science, it is necessary to deal with the large volume of data, and even more, data related to health. It should be noted that 43% of all daily bytes on the Web are related to Health and 47% of these are public health. In this way, the exhaustive volume of data requires organization and structuring for the possible support to the decision maker, which has led organizations to create search tools on specific websites on the Web, in order to extract and process the data to obtain essential information.

Free or paid access search tools are increasingly common and have an infinity in this competitive and / or collaborative intelligence, such as textmining, datamining software, etc. The practical result of the extracted and processed data generates essential information and is configured as useful, reasonable and strategic to streamline the decision process. Collaborative intelligence contributes strongly to the change of knowledge and power from the individual to the collective. The open source of collective intelligence will eventually generate results superior to the knowledge generated by proprietary software developed within corporations. Education and the way people are learning to participate in knowledge cultures outside formal learning contexts is crucial in the new global context. Learning through the means of collective intelligence is crucial, as it is important for the democratization of science, since it is interconnected with the culture based on knowledge and supported by the sharing of collective ideas, therefore, contributing to a better understanding of society’s diversity, with data open to institutions’ science. In this sense, new information science tools provide the facility, agility and synergy for collaborative work in a scientific and technological network, in such a way that the knowledge generated in the area can be explored and monitored for better action by decision makers; as in the case of epidemics and the COVID-19 pandemic.

COVID-19 GENERAL KNOWLEDGE

Large corporations also use the concept of open innovation as a way to provide innovation in their processes and / or products with contributions from any researcher, company etc. and in health, it also translates into an opportunity. Coronaviruses are a group of large, enveloped, positive-sense viruses and single-stranded RNA, belonging to the order Nidovirales, family Coronaviridae, subfamily Coronavirinae. At the end of 2019, a new type of coronavirus was discovered, provisionally called 2019-nCoV and later named SARS-CoV-2, due to its similarity to SARS-CoV. The disease caused by the virus was officially called Coronavirus Disease 2019 (COVID-19) by WHO. Two studies describe that by 2019, only six coronaviruses caused disease in humans: HCoV-229E, HCoV-OC43, HCoVNL63, HCoV-HKU1, coronavirus with severe acute respiratory syndrome (SARS-CoV) and coronavirus with respiratory virus in the Middle East.
Big data in health and open science: a contribution to the management of knowledge in COVID-19

(MERS-CoV). The first four are locally endemic and have been associated mainly with mild and limiting diseases, while the last two can cause serious diseases. SARS-CoV and MERS-CoV are betacoronaviruses and are among the pathogens included in the World Health Organization’s Model List of Priority Diseases.38

In order to understand the global scenario of COVID-19 cases, Johns Hopkins University provided the information dealt with Big Data in COVID-19, referring to the global epidemiological scenario (Figure 1) with 4,413,597 confirmed cases. As of May 14, 2020, Brazil had 196,375 confirmed official cases and a lethality of 13,555 cases.

Figure 1 – Confirmed cases of global COVID-19.

Brazil stands out in fifth place with 4.5% of deaths compared to the world ranking with a total of 300,315 deaths. The USA is the current epicenter of the pandemic with 1,405,961 confirmed cases and 85,066 deaths.

SCIENTIFIC KNOWLEDGE IN COVID-19

The biomedical literature base Medline, with more than 30 million citations in PubMed, has 36,687 scientific productions for the search term “Coronavir * or MERS-CoV * or SARS-CoV * or COVID * or HCoV or SARS or MERS”. Obviously extracting all documents from the base and studying them on time would take a long time to analyze. However, using search engineering algorithms and subsequent treatment of this volume, essential information can be obtained preliminarily to assist in knowledge management. For example, the free platform Carrot Search Lingo4G®.41 In this way, the contribution of information science to the evolution of the algorithmic basis is observed, where it enables faster and better processing, enabling the dissemination of scientific and technological knowledge produced.

With the same terms provided in PubMed, 100 most relevant documents were extracted. Of these, they were aggregated in sub-themes that correlate with new keywords that were most characterized (appeared) within each scientific production analyzed by the algorithm. As shown in Figure 2, in the central theme 23 documents were recovered (COVID-19 pandemic), in COVID-19 Patients 12 were separated, and for Coronavirus Disease 7 etc. It is worth mentioning that after data “mining”, the more “evident” terms in the search term searched, the new terms that are repeated more often are segregated and generate the cluster of the 100 most relevant in relation to the initial 36,687 in the PubMed database.
Regarding the number of authors, it was observed that there is participation in a varied network, from the opinion of one author to works with 11 (eleven) authors participating and with several international institutions.

**TECHNOLOGICAL KNOWLEDGE IN COVID-19**

Technological knowledge can be measured by the indicator of patents deposited in an industrial and / or intellectual property office in some country. Patents are a strong indicator of innovation in science and technology. Even with the extremely recent pandemic environment of COVID-19, one can observe the synergy of scientists and companies in the world, uniting the collaborative intelligence and open science, and that an exponential amount of scientific work has been produced, as seen previously, as well as in patents.

A challenge in the technological area is to combine the socialization of patent knowledge on an equal basis with scientists from any research group, even if there are no emergency situations of public calamity in the countries. According to the Intellectual Property Law (Nº. 9,279 / 96), a patent application is kept confidential for 18 (eighteen) months from the date of filing. Therefore, considering that COVID-19 was discovered in late 2019, it is to be expected that there will be no results for a patent. Thus, a search was performed with the synonyms of Coronavirus (family Coronaviridae; Severe Acute Respiratory Syndrome (SARS); Acute Middle East Respiratory Syndrome (MERS)). So, 03 (three) patent families were found: a Chinese CN110960532, by an independent inventor, a Korean KR10-2020-0032050 and an Indian IN202041010021.

In this sense, the search was carried out in an expanded way with other coronaviruses. Thus, Chart 1 shows 2,871 patent families found and distributed by the company holding the patent. Pharmaceutical company Pfizer stands out with 67 patents.
Big data in health and open science: a contribution to the management of knowledge in COVID-19

**Chart 1 – Major filers of patent families for Coronavirus.**

![Chart 1](chart1.png)

**Source:** Extracted by the authors of the European Patent Office by Questel Orbit®. May 2020.

Figure 3 shows the countries where patents were applied for by the filers of Chart 1 and / or the location of their inventors.

**Figure 3 – Countries of patent family filers for Coronavirus.**

![Figure 3](figure3.png)

**Source:** Extracted by the authors of the European Patent Office by Patent Inspiration®. May 2020.

In order to better stratify the knowledge generated and favor a quick understanding for the management of this knowledge, the essential data plotted on the map, assist the decision maker, in a practical observational understanding by the plotted figure. Therefore, in Figure 3, it is noted that the darker the color displayed in the country, it represents the country with the highest number of technological knowledge expressed in patent deposit. Note that the most accentuated color for red (USA) is where most.

**CONCLUSION**

The 21st century brought new challenges and great opportunities due to the growing volume of new data added to the Web on a daily basis. In Health, it is no different. Scientific and technological knowledge
has advanced exponentially never seen before. Thus, new ways to identify, extract and treat Big Data in health becomes fundamental for decision making.

Open science policies have been increasingly widespread in this new era of the 21st century, as a way of responding to this scientific and technological advance. In this sense, to help knowledge management, new solutions have been created, in order to speed up the arduous task of decision-making, especially in Health, especially in times of pandemic.

As a contribution to the management of knowledge acquired in record time of the pandemic COVID-19, the scientific ballast reached about 37 thousand publications in the PubMed database alone, 03 (three) families of patents and 2,871 family of patents related to the Coronavirus.

The results indicate that open science provides the synergy of results and networking. It is worth highlighting the limitation of “closed science”, for example, the result (knowledge) generated in the plot of Chart 1, was obtained by the Questel Orbit® information search and analysis system. Therefore, the organization that does not have the license will not be able to use it. However, it can be concluded that individual research has lost academic space in recent times, providing collaborative intelligence.

The scientific and technological data identified and the respective essential information made available, generating an essential framework for decision making, confirm the fundamental role of information science for all areas of Science - fundamental for Health. Digital convergence and synergy collaboration are preponderant factors for knowledge management in Health Big Data.

RESUMO

DESCRITORES
Saúde Pública; Big Data; Gestão do Conhecimento; Infecções por Coronavirus; Coronavirus.

REFERÊNCIAS


COLLABORATIONS
Jorge Magalhães (JM) and Zulmira Hartz (ZH) planned the article. JM and Gabriela de Paula Pereira developed the data collection together with Adelaine Antunes (AA) and the writing of the present work. Luc Quoniam and ZH analyzed, discussed the final revision of the work with JM and AA.

ACKNOWLEDGMENTS
Authors would like to thank the infrastructure of Farmanguinhos/FIOCRUZ and Gestec/FIOCRUZ for the availability of Questel Orbit databases, National Institute of Industrial Property (INPI) and Institute of Hygiene and Tropical Medicine (IHMT) at NOVA University Lisbon.

AVAILABILITY OF DATA
Not applicable.

FUNDING SOURCE
Own institutional funding.

CONFLICTS OF INTEREST
There are no conflicts of interest to declare.