

Factors influencing research data communication on Zika virus: a grounded theory

Michelli Pereira da Costa
michelli@unb.br

Fernando César Lima Leite
fernandoc@unb.br

Faculty of Information Science
University of Brasília, Brasília, Brazil.

Abstract

Purpose – The purpose of this paper is to propose a theoretical model to illustrate factors influencing research data communication on diseases caused by the Zika virus. **Design/methodology/approach** – The grounded theory methodology was adopted. For data collection, interviews were conducted with 13 Brazilian researchers involved with the Zika virus theme. Data analysis was performed using the open, axial and selective coding processes, according to the principles of the grounded theory. **Findings** – Based on data collection and analysis, seven theoretical categories representing intervening factors in Zika virus research data communication were identified. The emerging theory showed the centrality of researchers' expectations for acknowledgment and reward, influenced by the aspects of research funding and academic culture. Three factors involved in the macro processes of research data communication: data use, data production and data sharing. In the use and production processes, factors related to collaboration figured most strongly. In the production and sharing processes, the factors regarding data processing and the use of research data repositories were more pronounced. Finally, data sharing, and the possibility of reusing data are directly affected by the social context of Zika virus disease as an emerging disease. **Originality/value** – The study presents a theory developed systematically to explain the phenomenon of communication of research data on Zika virus. The theory presents a set of intervening factors of the process of communication of research data and discusses the factors in light of the fundamentals of information science.

Keywords: Research data, Scientific communication, Open data, Data sharing, Open science, Academic culture, Data communication.

Introduction

Deriving from Garvey's (1979) scholarly communication notion, which includes the ideas of production, dissemination, and use of information, the concept of research data communication in this paper refers to the processes of production, sharing, and use of data for the development of a given field of research. Initiatives to systematize such processes were documented even in the 1950s (Ruttenberg and Rishbeth, 1994), and in the 1970s, this topic was already found in information science literature (Søndergaard et al., 2003). In the 1990s, the discussion broadened from the Human Genome Project (HGP), which invited researchers from all around the world to share their data so as to contribute to the sequencing of the human genome (Zatz, 2000).

From the HGP experience, there has been a growth in the implementation of mandatory or recommendatory policies for research data communication, especially those resulting from public funding. The subliminal argument of these policies is that results from research funded by public resources is common property and, hence, should be widely accessible. However, the perspective of public science is also valid for research not funded by public resources, as research results should be broadly accessible in order to guarantee the characteristics of modern science as the historical construction of knowledge, assessment and refutation (Boulton, 2013).

The assumption with regard to science produced with public resources is that it should meet the needs of society as a whole and contribute to its social development. Two striking initiatives for data communication illustrated these aims. One of them was the aforementioned

HGP; as it summoned researchers to share human genome sequencing data, it underlined the importance of this action for the cure of illnesses affecting a great part of the global population, such as AIDS and tuberculosis. The other initiative was the research data communication policy of the National Institutes of Health (NIH) from the USA, which aims to optimize scientific development in the area of health sciences, thus contributing to human health improvement. These two proposals claim that benefits stemming from data communication will help improve collective health.

International initiatives in favor of open science and open access propose that wide access to scientific information, such as publications and original research data, can optimize scientific efforts and accelerate the advancement of science. A recent international initiative seems to confirm this expectation. In January 2016, the main research institutions and notable scientific publishers in the area of health sciences signed a declaration in which they committed to making scientific communication on the Zika and Ebola viruses freely available, in the form of data and publications (Dye et al., 2016).

The existence of the Zika virus has been recognized since 1947 in Africa. However, according to Vasconcelos (2015), it was only in 2004 that an epidemic resulting from the virus was registered in African and Asian countries. In 2015, the first reports of the virus in Brazil were established. By March 2016, 500 suspected cases of Zika virus disease had been registered, and three deaths were confirmed (Brasil, 2015a, b). In the same year, the European Centre for Disease Prevention and Control (ECDC) registered some cases of people infected by the virus in European countries (WHO, 2016), coinciding with a period of international interest in the virus.

In addition to being transmitted by the same vector that transmits the chikungunya virus, dengue virus, and yellow fever, the Zika virus possesses similarities to these diseases in its symptoms, means of prevention, and treatment (Chaves et al., 2015). What draws attention is that, long before the incidence of Zika virus disease, dengue fever was already a known disease in Brazil and other developing countries. In Brazil alone, 1.5m confirmed cases and 1,900 deaths from dengue fever were registered between 2013 and 2015 (Brazil, 2016). However, it was only after the incidence of Zika virus disease that the issue became an international concern. From such concerns, data communication actions and publications in the context of open science aimed to promote group efforts in overcoming the disease. Therefore, it became relevant to investigate the conditions that led the Zika virus to become an international concern and an object of open scientific collaboration.

From the scenario outlined here, the study on which this paper focuses sought to answer the following questions: from the researchers' point of view, what factors influenced research data communication about the diseases caused by the Zika virus? Why were research data on diseases caused by the Zika virus more likely to be communicated in the context of open science than other diseases? This paper aims to report on investigation results, and to propose a theoretical model to represent factors influencing research data communication about the Zika virus from the perspective of biological science and of health science researchers.

Methodology

This research adopted a systematized review and a grounded theory as its methodology. The systematized review of the literature was elaborated according to Grant and Booth (2009) proposal. Thirteen selected articles were selected from the Library and Information Science Abstracts (LISA) database (Table I). The search was conducted in mid-2016, through the search argument ((ti (“research data”) OR ti (“scientific data”)) AND sharing). The criterion used for the search privileged the documents that represent the subject of the sharing of the research data as the main subject and that reflected this subject in the title of the paper. The analysis systematized

from a form¹, allowed to identify a preliminary categories system about the theme. The Gephi software was used in the ordering of categories.

Table 1: Papers selected for systematized review

Database	Papers selected for systematized review
LISA search argument [(ti ("research data") OR ti ("scientific data")) AND sharing]. mid 2016	Sturges et al. (2015), Higman and Pinfield (2015), Whitmire et al. (2015), Knight (2015), Thompson et al. (2014), Melero and Hernández-San-Miguel (2014), Douglass et al. (2014), Dallmeier-Tiessen et al. (2014), Bond et al. (2013), Cox et al. (2012), Losoff (2009), Childs and Mcleod (2004) and Gardner (2004).

Source: Elaborated by the authors.

According to Strauss and Corbin (2008), grounded theory is a research strategy that allows the proposition of an explanatory thesis of a phenomenon based on gathering and organizing analysis from researched data; therefore, the theory emerges from data analysis. The main data collection resource to elaborate a grounded theory is the conducting of interviews with a theoretical sample. This sample concerns essential subjects for the development of the questions to be studied.

In order to answer the research questions, Brazilian researchers directly involved with research on “Zika virus” and “Aedes aegypti”² became the universe of subjects in research participating. Researchers were identified by means of research group registers in the Group Directory of Research in Brazil (DGP/CNPq), from the search criteria described in Table II. The use of the strategy described in Table II resulted in the identification of 13 groups for research aspects directly concerning the Zika virus and Aedes aegypti. All groups were contacted and invited to participate in the survey, however, only 10 groups responded to the invitation within the stipulated period. Therefore, this was the intentional sample selected for the universe of Brazilian researchers, which Strauss and Corbin (2008) call a homogeneous theoretical sample. Its homogeneity represents the singularity³ of the group of interviewed subjects, apparently showing the same level of knowledge and involvement with the studied phenomenon. On the other hand, the theoretical character of the sample points to the fact that the elements investigated in the interviews are theoretical categories and do not seek to represent the plurality of opinions of all subjects involved with the topic. The interviews were carried out with the coordinators or leaders of the selected groups with the help of a semi-structured script⁴. All interviews took place between the middle of 2016 and the beginning of 2017.

Table 2: Research group search strategy

Step	Strategy	Result
1	Parameterized query. Search word: “aedes aegypti” in the fields “Research Group name” and “Research line name” + filter – certified groups and updated groups	10 registers
2	Parameterized query. Search word: “Zika” in the fields “Research Group name” and “research line name”	5 registers
3	Addition of results obtained for search word “aedes aegypti” and the results for “Zika” (results 1 and 2)	15 registers
4	Removal of two repeated groups in the first and second search	13 registers

Source: Elaborated by the authors.

¹ The analysis form used in this step was presented in the full research report, p. 94, available at: <http://repositorio.unb.br/handle/10482/23000>.

² It is important to include the term aedes aegypti once it concerns the transmission vector of Zika virus.

³ In Portuguese: Diretório de Grupos de Pesquisa no Brasil. Source: <http://lattes.cnpq.br/web/dgp>.

⁴ The interview’s semi-structured script is available on the theses reporting the entire research. Source: <http://repositorio.unb.br/handle/10482/23000>.

The interviewed researchers varied in their regions, age, gender and area of knowledge (Table III). In the interviews, researchers were led to talk about the phenomenon of open science, their motivation to produce and share their data and also to reuse data from other researchers. Furthermore, the researchers were prompted to identify particularities involving research on the Zika virus.

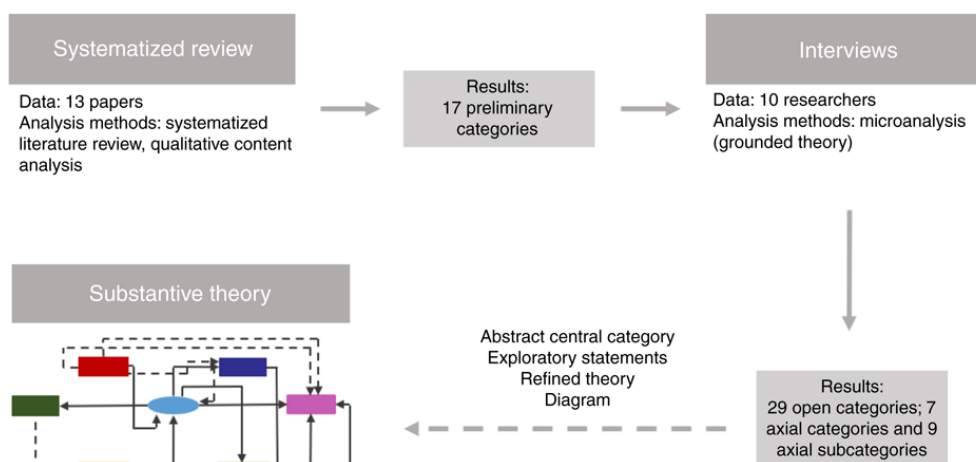
Table 3: Characterization of the interviewed researchers

Researcher	Area	Region	Age	Gender
P01	Biology	Midwestern Brazil	30-40	Male
P02	Biology	Southeastern Brazil	40-50	Female
P03	Pharmacy	Southeastern Brazil	40-50	Female
P04	Pharmacy	Southeastern Brazil	30-40	Female
P05	Biology	Southeastern Brazil	50-60	Male
P06	Biology	Southeastern Brazil	30-40	Male
P07	Biology	Midwestern Brazil	40-50	Male
P08	Medicine	Southeastern Brazil	50-60	Female
P09	Medicine	Midwestern Brazil	70-80	Male
P10	Pharmacy	Southeastern Brazil	30-40	Male

Source: Elaborated by the authors.

The proposition of a theory concerning the phenomenon was developed from the microanalysis of the data gathered in the interview. In short, such data underwent three examination processes that made it possible to encode the data and theoretical comparisons. The first was the open encoding of data, which aimed to develop or enhance the categories according to their characteristics and attributes. Thus, the phenomenon was labeled by the categories and subcategories emerging from the data. The following step was axial encoding, in which a new assembly of the categories and subcategories took place according to a central category. The last microanalysis step was selective encoding, which aimed to refine the theory, seeking a logical and internal consistency. In this step, the relations between the abstract central category and the other categories and subcategories were deepened by means of exploratory statements. According to Strauss and Corbin's (2008) proposal, the presentation of the refined theory should include variations of the categories and be accompanied by a diagram, as is featured later in this paper (Figure 1).

Figure 1: Research methodology

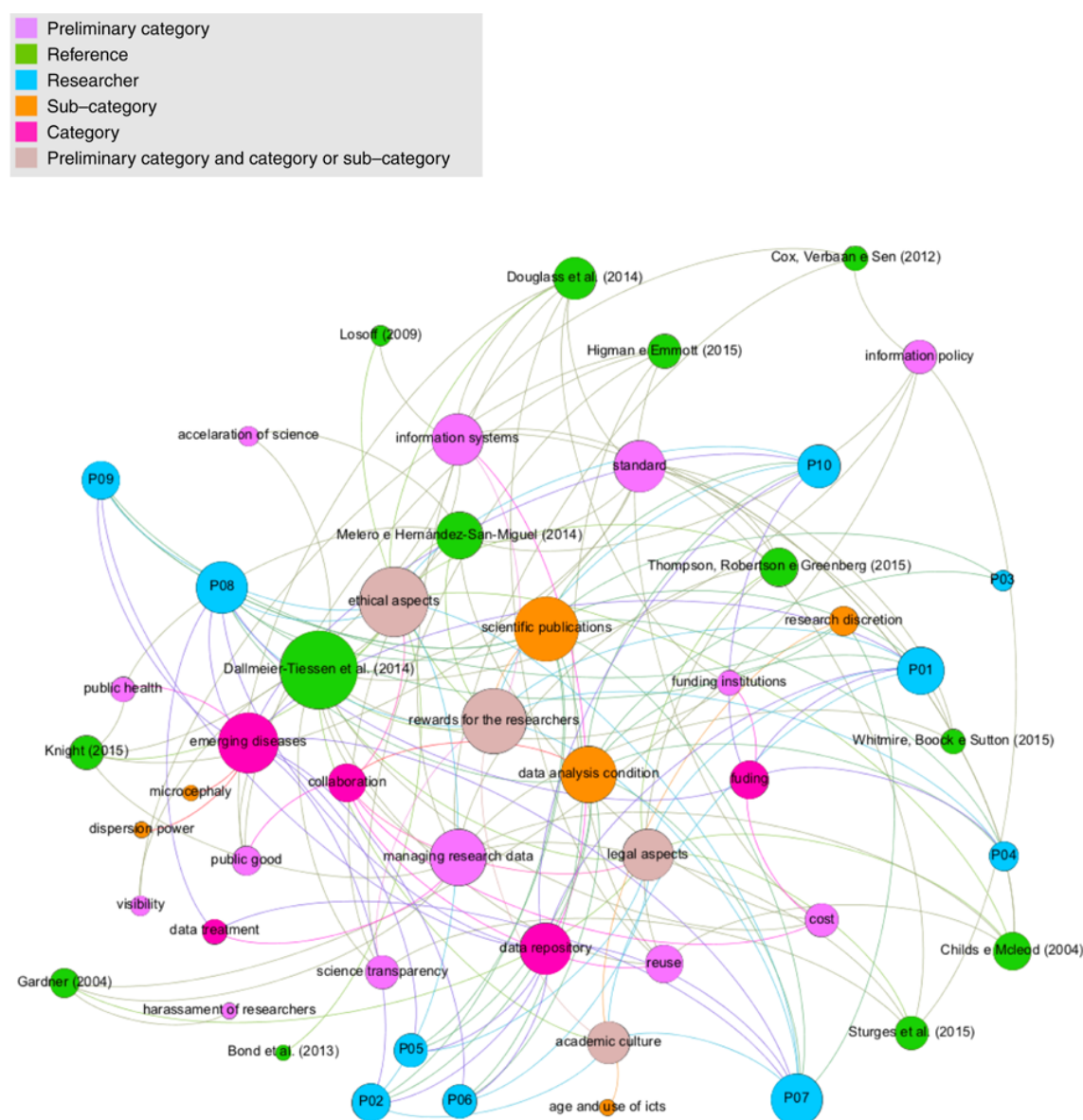


Source: Elaborated by the authors

Factors that influence research data communication, according to information science literature

Scientific literature in the field of information science took a role of preliminary source of information on the factors, which influence communication of research data. The systematized review of the literature allowed the identification of 17 factors that influence the processes of communication of research data. The result was obtained by performing three steps. The first consisted in the free labeling of the elements. In the second step, all elements were translated into similar terms called preliminary categories. In the third step, the Gephi software was used for to connect the categories and the authors (Figure 2).

Figure 2: Preliminary categories



Source: Elaborated by the authors with Gephi Software

The figure is the result of a network analysis methodology, as explained by Jacomy et al. (2014). The network represents the relation of the authors and the intervening factors mentioned by them. Each article, represented by its authorship, and each factor represents a node in the network. The nodes closest to the center of the network represent the authors and concepts most

relevant to the proposed analysis. Therefore, the relevance of the discussion proposed by Dallmeier-Tiessen et al. (2014) and Melero and Hernández-San-Miguel (2014).

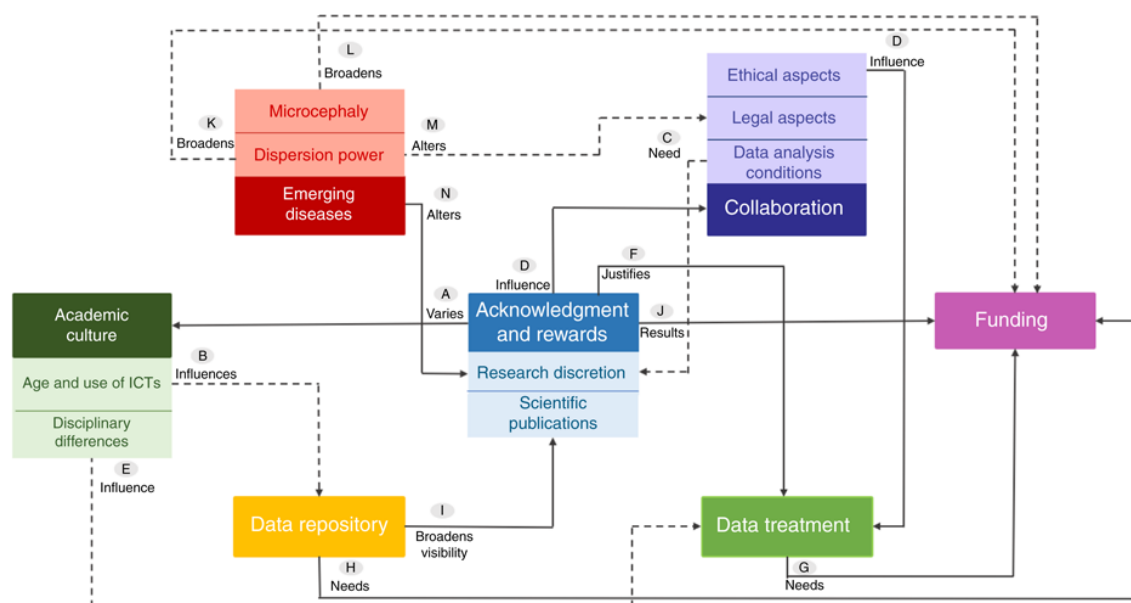
In addition, it was possible to organize the factors from the analysis dimensions. The organization based on conceptual approaches between factors and results in the proposition of three dimensions. The first concerns the imperatives of modern science, where nine categories are identified that relate conceptually to the prerogatives of modern science: accelerating science, reusing research data, science transparency, harassing researchers, visibility, rewarding researchers, academic culture, legal aspects and ethical aspects. The second dimension deals with aspects related to infrastructure, whose recurrent categories were information systems, standards and policies. Finally, the funding dimension of the research data communication process emerged as prominent, in which the categories of research funding institutions, research and data as a public good and associated costs were highlighted.

In this paper, the preliminary categories will be discussed based on the results obtained in the interviews, in the next section.

Grounded theory on factors influencing research data communication on Zika virus

Research data communication has been pointed to as a phenomenon capable of improving the flow of communication and optimizing the processes of science, as Gezelter (2009), Lopez (2015) and Hey and Trefethen (2008) indicate. Despite its benefits, open communication of data is still not the reality of research practice and, when it does take place, it is expressed with singularities. In the context of international guidelines for open communication of publications and data concerning the Zika virus theme, identifying the factors influencing this process becomes appropriate. From the methodology described, seven factors and their respective subfactors were identified that preponderantly influence research data communication on the Zika virus. Figure 3 illustrates the grounded theory resulting from this research, in which factors, subfactors, and their relations are presented. The relations are indicated by arrows, verbs, and letters from A to N, in a scheme guiding the explanation of the following theory.

Figure 3: Diagram of the substantive theory on factors influencing research data communication on the Zika virus



Source: Elaborated by the authors

Acknowledgment and rewards

Every researcher discussed based on factors in order to establish the acknowledgment and rewards category. The main motivation for conducting studies on the Zika virus, as well as for divulging their results, is the expectation researchers have to be acknowledged and rewarded for their work. These arguments related to systematized factors in the preliminary categories of science acceleration, academic culture rewards given to researchers and information system, based on ten of the 13 papers analyzed.

Papers and all interviewees pointed out that academic publications are the most essential expression of acknowledgment. They show prestige within the academic environment and, despite variations resulting from disciplinary differences, they result from the academic culture of modern science. However, the communication of research data does not receive as much academic prestige as papers. According to Dallmeier-Tiessen et al. (2014), researchers argued that sharing research data does not generate status or promotion in their careers, so it did not compensate for their efforts. This discussion was guided by Melero and Hernández-San-Miguel (2014), Dallmeier-Tiessen et al. (2014) and Losoff (2009). In general, the authors emphasize that the low academic value that is attributed to the disclosure of the data added to the lack of incentives for the practice are aspects that inhibit the process of dissemination of research data.

In a research published in 2011, Whyte and Pryor infer the need for acknowledgment and rewards. In the study, the authors indicated that citation methods of the research data could be an alternative to give recognition to the researchers who collected the data. The alternative is related with the Brase et al. (2009) proposal about the citation methods for the context of the research data. Despite the alternative pointed out in the literature, the researchers interviewed by this study did not recognize the citation of the research data as compensatory for the effort involved in collecting the data. For them, what is able to confer such reward is the papers.

The three investigated areas of knowledge: biology, medical science, and pharmaceutical showed dependence relation between acknowledgment and journals. The expectation in divulging research results in journals points to another factor intervening in the process, treated here as the subcategory “research discretion.” Discretion, hence, is seen as a requirement for publication in journals, which, as a rule, require a work to be original and not previously published. Therefore, discretion hinders the communication of open data. This obstacle is widely acknowledged and discussed in scientific literature on the theme and, according to Walport and Brest (2011), the intention of maximizing the quantity of scientific publications limits the possibility of using data. Even if the full set of data is only available after the publication of the article, some researchers report damage to their work. However, in the field of Biology, the requirements to make the data available in proper repositories have been presented repeatedly as a condition to publishing papers. Yet, in the field of pharmacy, the process seems to operate conversely to open science tendencies. Comparatively, researchers in the field of pharmacy were shown to be less likely to open a data set to the public at the moment of publishing or after a period of exclusive exploration, once it was able to originate new publications and, especially, new technologies. In the case of technologies, information resulting from research, as well as its data, are communicated in the form of a patent, safeguarding the exclusive exploration of its content for the author or whoever holds the patrimonial rights.

The need for discretion about research data is so high in the field of pharmacy that even the practices of data registry are adequate to ensure discretion. According to the report given by P03, the data considered most promising are registered in codes, so that other researchers are unable to comprehend them. Therefore, precautions for research discretion begin with the process of data production, and are reflected in the steps of sharing and reusing data (or the absence of these).

In general, research discretion was shown as a necessary behavior for scientific practices, especially those bringing acknowledgment and reward to researchers for their work. The intensity with which the discretion is desired varied between the areas of knowledge.

Academic culture

The central and root element of the investigated phenomenon revealed that researchers have interest in being acknowledged and rewarded for their research efforts. However, it was notable that although being acknowledged is the main motivation for researchers to divulge their research results, the inclusion of the full data set in the group of products resulting from the research varies according to the academic culture (A). Academic culture concerns the elements affecting the researchers' behavior, as Dallmeier-Tiessen et al. (2014) defined it. According to the authors, cultural differences between the areas of knowledge are especially due to the nature of the data and how researchers communicate their research results.

Researchers P01, P02, P05⁵, P06, and P07 are essentially related to the field of Biology. Compared to the other two areas, this area appeared much more inclined to research data open communication practices. The data from these studies concern mostly molecular genetics, resulting in the systemization of databases. Even in cases when information from patients is collected, the resulting research data concerns molecular genetics. Such data present the following characteristics, differentiating them from other data: they do not identify the patients; they are disclosed with no need for analysis; they are applicable to other areas of research. In general, the data are made available in data repositories such as genetic databases.

Genetic databases, or genetic data repositories, have been developing and gaining strength in discussions about open science since the time of HGP in the 1990s, and afterward with NHI's information policy (National Institutes of Health, 2003). Therefore, historicity points to genetic data repositories as one of the first international initiatives favorable to open science, contributing to an understanding of factors, and positioning biology as an area of knowledge closer to open data practices.

According to Whyte and Pryor (2011), the life science, chemistry and astronomy are fields more "open working." However, this result shows the biology is the life science more open working than medicine and pharmacy in the study context.

Researchers P08 and P09 work with researchers in the field of medicine. Their researchers depend directly on patient information. Because of their typology, these data can require or result in the identification of the patient, and often need to be contextualized before new analyses can be done. The data are disseminated, necessarily, in journal articles or other traditional scientific publications requiring previous data selection. Therefore, it was verified that, for the field of medicine, there are no information systems to disclose raw research data.

Researchers P03, P04, and P10 perform research in the area of pharmacy, aiming to develop a vaccine for the Zika virus. The data needed to develop such research include those pointed out by the area of biology, and that resulting from experiments with chemical compounds. In both cases, the data differ from that in the field of medicine, as they do not deal directly with patient information and do not require contextualization for analysis. Furthermore, they differ in the nature of the results they generate or aim to produce. The results are intended to generate a patentable and/or marketable product, hence the research processes and results require more discretion from the researchers. From the researchers' reports, it was possible to observe that such aims affect all data communication macro processes. This relation had already been reported by Childs and Mcleod (2004) when they pointed out that intellectual property rights, by means of patents, regulate both storing and access practices and the reuse of protected data. According to Childs and Mcleod (2004) such rights can regulate the practices of storage, access and reuse of research data. The authors did not explore the nuances of the relationship, they just pointed it out. In the same sense Sturges et al. (2015), Gardner (2004) and Thompson et al. (2014) and Dallmeier-Tiessen et al. (2014) presented the theme as one of the barriers to the wide communication of research data.

⁵ Researcher P05 has a Medical degree, however, P05 is directly involved in the field of Biology.

In general, it was determined that the types of data involved in research about the topic are quantitative in nature, with genetic databases and images, and qualitative, with texts. Such results are close to those presented by Whitmire et al. (2015) for the area of public health. The channels of data disclosure vary according to the nature of the data and the area of knowledge, as, for example, data repositories for biology and patents for Pharmacy.

Disciplinary differences are an explanatory element of academic culture expressed in practices. As Whitmire (2002) defined, disciplinary differences are differences in characteristics, norms, and values among areas of knowledge.

Researchers in the area of biology claimed that, within their field of study, the age of the researchers involved influences their willingness to openly communicate their data and publications (B). In this sense, younger researchers were shown to be the most prone to the practice, both due to their abilities in the use of technologies and to their interest in collaborating with other research projects. The same result was observed in the study by Dallmeier-Tiessen et al. (2014), which found that the use of technologies for open research data communication, such as data repositories, is more frequent among young researchers.

Brazilian researchers and by Dallmeier-Tiessen et al. (2014) pointed a tendency that does not confirm the results of the Fecher et al. (2015) study. The authors conducted a survey about the reuse of research data among German researchers and found that researchers over age 50 are more likely to make their data available to others than younger researchers. According to the authors “this result resonates with an assumed influence of seniority in the academic system and competitiveness on data sharing behavior.” The relationship between the age of the researcher and the tendency to open data in the Brazilian context is different from the German context described by Fecher et al. (2015). However, both results indicate that the sharing of research data does not confer the career rewards of researchers, such as papers.

Collaboration

Every researcher discussed based on the factors in order to establish the collaboration category. The aspects indicate the collaboration as necessary to comply with legal and ethical requirements of research. The perspective of collaboration was also identified as an alternative to overcome the problems of data reuse and the limitations of research funding. These arguments are related with the systematized factors in the preliminary categories of ethical and legal aspects, reuse, cost and public good, based on the Dallmeier-Tiessen et al. (2014), Douglass et al. (2014), Melero and Hernández-San-Miguel (2014), Bond et al. (2013), Childs and Mcleod (2004), Thompson et al. (2014) and Knight (2015).

Research discretion was indicated as a condition for publishing in a scientific journal. This is because it is common for publishers to require that articles present original, unpublished results. In addition to the interest in publishing, researchers justified the need for their work not to have been previously published due to the conditions available for data analysis (C). Researchers in the areas of biology and pharmacy think they lack the necessary infrastructure for quick data analysis, as do researchers in developed countries. Therefore, the results indicate open research data communication would be to the particular disadvantage of researchers in developing countries.

To elaborate further on this question, it is important to consider that which places researchers from developing countries at a disadvantage in relation to data openness is their precarious access to equipment and other materials needed for analysis, such as reagents. However, in the context of Zika virus research, the interviewees were of the opinion that, compared to researchers in developed countries, they experienced greater ease in obtaining biological samples, which is a type of resource needed to develop research. From the researchers' discourse, it was observed that the biological material from patients infected with the Zika virus has become an important element in balancing the competition for Brazilian researchers.

The advantage brought about by the abundance of biological material available to Brazilian researchers still required the help of a legal instrument: Law No. 13.123/2015 (Brasil, 2015b). This

law aimed to regulate the access of foreigners to what it calls genetic patrimony (Santilli, 2015). Thus, it became more difficult to send biological samples abroad. Such restriction has led researchers from other countries to seek access to biological material by means of a collaboration with Brazilian researchers (D), intensifying collaborative research on the Zika virus.

In this scenario, the disadvantages in data analysis and the advantages in access to biological samples for Brazilian researchers became an important element to justify the collaboration between Brazilian and foreign researchers. While Brazilian researchers need more and better resources to analyze the data, foreign researchers need biological samples to generate data. However, despite the reciprocal dependence, data analysis is more valued by science than data collection, as the interviewees stated. As a rule, data analyses are published in scientific articles, which results in acknowledgment and rewards. The collection and processing of data do not generate the acknowledgment that researchers desire, and constitute merely a means to perform the analysis and publish their results.

This concern expressed by the interviewees was not identified in any of the articles analyzed in the systematized review. However, was a theme for international debate published in the journal *The Lancet*, Nos 377 and 378. In the first publication, Walport and Brest (2011), managers for research agencies Wellcome Trust and Hewlett Foundation, defend a balance of rights and responsibilities for those who generate and use the data. Although they consider the problem, the authors do not elaborate on what a fair use of data would be; they only state that, as it concerns public health, the data should receive maximum use. In response to the arguments presented by Walport and Brest (2011), Sankoh and Ijsselmuiden (2011)⁶, researchers from Africa, state that fair use should guarantee all researchers equal infrastructure and resources to analyze the data. Therefore, only this parity in means and capacities of data analysis could confer justice in the use of the collected data.

In addition to the necessary conditions for and legal aspects of data analysis, another element elucidating collaboration are the ethical aspects of research (D). In scientific literature on the theme, one of the main ethical concerns is participants' privacy, which is, in general, considered a hindrance for open data communication (Dallmeier-Tiessen et al., 2014; Douglass et al., 2014; Melero and Hernández-San-Miguel, 2014; Bond et al., 2013; Childs and Mcleod, 2004; Thompson et al., 2014; Knight, 2015). Despite the understanding that participants' privacy is a sovereign right, Walport and Brest (2011) state that privacy can be guaranteed by means of data anonymity, thus guaranteeing maximum data use. For these authors, broadening data use allows more results to be obtained with no need to collect the same data again; it is, hence, also an ethical principle to be respected in public health.

According to interviewees' perception, a commitment to the preservation of participants' privacy and dignity requires that the data be disclosed in such a way not to identify them. However, the removal or effacement of data can affect the usability of the data set by researchers external to the original research. The concern with the ability to understand the data was also signaled by Fecher et al. (2015) study and was pointed as an factor influencing research data communication.

Biological material is used to generate data in all three areas of knowledge identified, but only the area of medicine deals directly with information about and contact with the patients. Therefore, this area is more involved with issues concerning ethics in research than the others. Especially in the field of medicine, it has been suggested that making data available for the use in other research implies the participation of at least one researcher who took part in the gathering and analyzing of data for the original research.

Data processing

The factors discussed by researches P07 and P08 were the basis for the data processing category. The aspects indicate the need for standardization of data, appropriate professionals and

⁶ Published in medical journal, *The Lancet*, issue 378.

specific funding for the management of research data. The researchers highlighted that the processing of the data in order to share it requires more efforts, therefore, more resources are needed. These arguments are related with the systematized factors in the preliminary categories of managing research data based on the Dallmeier-Tiessen et al. (2014), Thompson et al. (2014) and Douglass et al. (2014).

The difficulty in reusing data is due to the need to understand the context of the original research and, especially, how the data were processed. This means information referring to the context, to how the data were collected, and the variance presented. Even observations about the research environment and participating subjects are necessary for other researchers to reuse the data from the field of medicine. Therefore, data processing was considered as another explanatory category for the phenomenon of research data communication involved in disciplinary differences (E).

For studies in the healthcare field, a data registry aiming for full disclosure implies in standardized values, so individuals will understand the data despite its anonymity. In the field of pharmacy, the main challenge concerning data processing has to do with storing the complete data set in digital format. The use of personal computers is not allowed in the laboratories where the experiments are performed. Hence, all the data are written by hand to be transcribed later. The data to be collected pertains only to the interests of the research to which it is strictly related, disregarding, thus, its possible reuses. Data processing practices in biology research were more favorable to open communication. The data are originally registered in a computer, under internationally established standards. The use of English represents an additional effort required for open communication in the process of data registry. The interviewees' opinion was that data processing is a hindrance to open communication, which seems to be in accordance with issues discussed by Dallmeier-Tiessen et al. (2014), Thompson et al. (2014), and Douglass et al. (2014). According to the authors, data processing for open communication requires more time and effort, as it is necessary to structure the data according to a given standard, or describe the data more thoroughly than the original research would require.

We conclude, hence, that data processing intended for open communication requires more effort than what would be needed for the researchers' private use. Although there are barriers in this process, two factors offer compensation for the effort. The first concerns the possibility of publishing the research results in journals of great international prestige. For this, researchers tend to adjust to the publishing policies of such journals, and will alter the form of the data registry if necessary (F).

Funding

The factors discussed by researchers P01, P02, P04, P08 and P10 were basis for the funding category. The aspects indicate the need to share responsibility with institutions interested in the treatment and availability of data were raised. The arguments also pointed out that adequate funding for the treatment and availability of data can generate savings in resources in future research funding. These arguments are related with the systematized factors in the preliminary categories of financing institutions and costs, based on the Dallmeier-Tiessen et al. (2014), Douglass et al. (2014), Melero and Hernández-San-Miguel (2014) and Higman and Pinfield (2015).

Factors related to funding highlight research funding agencies and universities as strategic agents in promoting the open communication of data. In the discussion by Dallmeier-Tiessen et al. (2014) and Douglass et al. (2014), research funding agencies hold the authority to establishing parameters for collecting, divulging, and using the data from the research they fund. In short, the authors state that this determination must be made via information policies, wherein the requirements for producing, divulging and (re)using data should be established. The interviewees confirmed the projection of authority on funding institutions for research. Funding would then be directed to the necessary resources for data processing in the form of information systems and services. Two aspects of services were mentioned. The first was the hiring of data specialists. The

second aspect regards appropriate information systems, such as software and equipment that reduces the effort required by researchers in producing and disclosing data.

The relevance of the financing to the data treatment was also pointed out by Whyte and Pryor (2011) as a factor influencing in the communication of the research data. According the authors, the researchers recognize the benefits of communicating research data, especially to increase the visibility of their research and that the reuse of data generates resource savings. However, open communication of data requires more efforts in data processing and therefore adequate funding is required to ensure the practice.

From the interviewees' perspective, scientific journals are the main means of acknowledgment for their work. Such acknowledgment brings about two desired results: academic status, which varies according to the academic culture, and the broadening of the possibilities for research funding (J). Funding was presented as a means to obtain the infrastructure needed to develop the research, and as a result, a means for compensation. Therefore, funding represents both a means for open communication of data and the goal of the process. Thus, institutions that funding research, as well as scientific journals, are determining factors in the process of open communication of research data.

Data repositories

The factors discussed by researchers P01, P02, P05, P06 and P07 were basis to establish funding category. The aspects indicate the database as infrastructure required for the management and sharing of search data. These arguments are related with the systematized factors in the preliminary category of information system, based on the Losoff (2009), Sturges et al. (2015), Childs and Mcleod (2004) and Dallmeier-Tiessen et al. (2014). In synthesis, these authors considered data repositories as efficient alternatives to managing research data, as they can absorb the diversity in volume and types of data files the research may produce.

In addition to offering services, such repositories are information systems responsible for storing, organizing, disseminating and preserving research data. These services assume mediation by information professionals who, once linked to the institution, represent its interest and institutional commitment in treating and disclosing the research data it funded. While the commitment is established, research funding agencies can use repositories as a tool to control the research outputs of what they funded.

This perspective was discussed by researchers from the field of biology. For them research data repositories (H) as necessary information systems to make the open communication of data possible. It is relevant to point out that the discussions about research data repositories were at the level of ideas, as none of the interviewed researchers discussed this type of system as offered by an institution. Nevertheless, repositories were repeatedly mentioned as necessary to operationalize research data open communication.

The use of research data repositories supports the quest for acknowledgment and rewards in two senses (I). The first is in the sense of systematically broadening the sharing of results. Especially in areas such as biology, repositories are expected to reduce the effort required for data registry and, simultaneously, generate more visibility for the research and for the institutions where the research was conducted. The visibility attained by managing the data by means of open repositories is a necessary condition to obtaining more acknowledgment and, later on, may result in rewards. The second sense regards visibility for research funding. Thus, research data repositories could be tools to measure the accountability of the funded research, in addition to helping in the assessment of the obtained results.

Emerging diseases: the case of the Zika virus

Every interviewee mentioned Zika virus subject as another intervening factor in research data communication practice. The singularities of this theme resulted in greater offers for funding

and new conditions for conducting research. The funding for studies on the Zika virus is more generous than that given to other emerging diseases. The interviewees reported that their research on the topic received more funding when compared to research on other topics. The increase in the quantity of available resources was attributed to Brazilian public research funding institutions, and to the singular interest of rich countries in this topic. In 2016, the Brazilian Government made about R \$550,000,000.00⁷ available for research related to the Zika virus. Furthermore, resources were made available by the World Health Organization (WHO), the European Union, and the USA (WHO, 2017).

The interest in the topic and the amount of funding for research were attributed especially to the spread of the virus throughout the world – a factor substantially different from other epidemics of emerging diseases in developing countries (K). The world threat of the virus was announced when the possibility of its sexual transmission was verified, as Musso et al. (2015) documented. Thence, the spread of the virus ceased to relate to climatic and sanitary issues, typical of developing countries. The global spread of the virus was attributed to the International Federation of Association Football (FIFA) World Cup that took place in Brazil in 2014. On this occasion, Brazil received a large number of tourists from many nations. After the event, cases of European and North American patients infected by the virus were registered (Vasconcelos, 2015; Zanluca et al., 2015).

The spread of the virus attracted new research funds, and the interest of the pharmaceutical industry. Therefore, the incidence rate of the virus in patients from developed countries, in addition to funds from agencies in developed countries, attracted the attention of the pharmaceutical industry and large corporations. The atypical interest in the Zika virus is relevant for discussion, as other diseases of similar nature – some even considered more serious than Zika – are not an object of interest for the pharmaceutical industry. Dengue, chikungunya, and yellow fever are examples of diseases transmitted by the same vector as Zika, and, according to the interviewed researchers, may result in more serious consequences than Zika. Nevertheless, despite receiving funds from Brazilian governmental agencies, such diseases did not attract any interest from the pharmaceutical industry, nor the same attention in developed countries.

Investigations into the Zika virus gained visibility from studies performed in Northeastern Brazil. Its international reach was due to its high power of dispersion and the confirmation of cases of microcephaly in newborns, as pointed out by Mlakar et al. (2016), a phenomenon the present study confirms (L). Therefore, from the interviewees' perspective, these two factors were determinants in the international visibility of this theme.

These two aspects, the microcephaly and the spread of the virus, in addition to the epidemic process the virus brought about, resulted in the classification of the epidemic as an international public health emergency. Such classification alters the availability of research data and results, thus creating exceptions to the advantage of open communication of research results. According to the news article Associated Press (2016), the interest in accessing data collected in developing countries was one of the reasons that led WHO to declare the epidemiological problems of the Zika virus an international emergency.

As Carmo et al. (2008) explained, the classification of an epidemic as a public emergency originated in 1992 in a proposition by the Institute of Medicine, also known as the National Academy of Medicine (NAM), in the USA. This classification aimed to create an international alert for infectious diseases with increasing incidence among human beings. The definition proposed in 1992 underwent alterations until 2005 when it was established in the International Health Regulations (IHR). The classification of an epidemic applies to diseases posing public health risk to other countries, which then requires a coordinated international response (WHO, 2009).

⁷ Approximately \$150m, as of 2016 average conversion values. The Brazilian Federal Government announced value results and amounts at the Federal Senate webpage (Source: www2.senado.leg.br/bdsf/bitstream/handle/id/518252/noticia.html?sequence=1) and by Financiadora de Estudos e Pesquisas (FINEP) at its webpage (Source: www.finep.gov.br/noticias/todas-noticias/5223-finep-vai-investir-r-230-milhoes-contra-o-zika-em-2016).

Coordinated actions are performed to monitor the rate of incidence, and to control the transmission vector. Such actions are determined by World Health Observances in accordance with IHR and affect all WHO member nations. Hence, according to Bruniera-Oliveira et al. (2014), the classification of an epidemic as an international public health emergency implies four actions:

- report all cases to WHO;
- implement all WHO recommendations;
- develop national and regional capacity for epidemiological observation; and
- share data in national and international ambit in real time.

The classification of an epidemic as a public health emergency commits WHO member nations to disseminate all data regarding the incidence of the disease in their territories, as well as any necessary data to control the phenomenon. The mandatory feature of information disclosure further determines it should be made in real time, thus affecting practices and interests so far discussed by researchers and the academic culture as a whole. This influence was confirmed by the interviewed researchers from all three areas, although the emphasis was employed specifically for the phenomenon of the Zika virus, and did not include all international public health emergency themes.

Even in the area of pharmacy, which proved to be the least likely area for research data open communication, behaviors more favorable to open science were found in the specific context of the Zika virus. According to P10's report, studies regarding this theme were the most progressive regarding openness, reach and rapidity of result sharing.

The classification of a disease as an international public health emergency is made by WHO and requires joint actions from all affected member nations. One of these actions regards disclosing information on incidents, as well as making data available, including biological samples. At the time preceding classification by the WHO, Brazilian research institutions were involved in controversy with foreign researchers regarding the impossibility of sending data abroad. The difficulty in sending data, especially biological samples, was due to Law No. 13.123/2015, which treats biological samples collected in Brazil as national genetic patrimony, thus guaranteeing their author or rights holder exclusive exploration of the material. However, the right is not absolute and can be revised in the case of a collective right. Therefore, the understanding of diseases caused by the Zika virus as an international emergency is a factor in removing the legal barrier that prevents sending data to foreign researchers, thus widening communication, as indicated in the results of this study (M).

Conclusions

A selective encoding strategy elaborated the theoretical model representing factors influencing research data communication, and leading to the proposition of a substantive theory. The theory, hence, indicates relations among the categories and subcategories in the process of explaining the studied phenomenon.

The theory emerging from the data indicates the centrality of the expectation for acknowledgment and rewards in the process of communicating data. Such expectation, affected by aspects concerning research funds and academic culture, influences the three macro processes of research data communication. Among the processes of using and producing research data, factors regarding collaboration stand out, especially those concerning data analysis conditions. On the other hand, among the processes of producing and sharing data, factors regarding data processing and the use of research data repositories and information systems of this nature were observed. Finally, sharing and the possibility of data reuse are directly affected by the social context of emerging diseases. Such context promotes legal and ethical conditions and the necessary infrastructure to favor open data communication.

The study produced a substantive theory about the factors influencing research data communication on the Zika virus. Such factors indicated both favorable conditions for and barriers

to the practice of open communication. Therefore, the produced theory constitutes an analytical tool to comprehend reality, and allows guidelines for intervention in research data communication practices, aiming to promote open science.

References

- Associated Press (2016), “G1– Brasil sonega amostras de zika para pesquisa no exterior, dizem cientistas – Notícias em Ciência e Saúde”, available at: <http://g1.globo.com/ciencia-e-saude/noticia/2016/02/brasil-sonega-amostras-de-zika-para-pesquisa-no-exterior-dizem-cientistas.html> (accessed March 30, 2016).
- Bond, C.S., Ahmed, O.H., Hind, M., Thomas, B. and Hewitt-Taylor, J. (2013), “The conceptual and practical ethical dilemmas of using health discussion board posts as research data”, *Journal of Medical Internet Research*, Vol. 15 No. 6, p. e112, doi: [10.2196/jmir.2435](https://doi.org/10.2196/jmir.2435).
- Boulton, G. (2013), “Reinventing open science for the 21st century”, in Rodrigues, E., Swan, A. and baptista, A.A. (Eds), *Uma década de acesso aberto na UMinho e no mundo*, Braga: Universidade do Minho, Serviços de Documentação, Lisbon, pp. 239-250.
- Brase, J., Farquhar, A., Gastl, A., Gruttemeier, H., Heijne, M., Heller, A., Piguet, A., et al. (2009), “Approach for a joint global registration agency for research data”, *Information Services and Use*, Vol. 29 No. 1.
- Brasil (2015a), Protocolo de vigilância e resposta à ocorrência de microcefalia, Ministério da Saúde, Brasil, available at: <http://portalsaude.saude.gov.br/images/pdf/2016/janeiro/22/microcefalia-protocolo-de-vigilancia-e-resposta-v1-3-22jan2016.pdf>
- Brasil (2015b), “Lei no 13.123/2015 Brasil”, available at: www.planalto.gov.br/ccivil_03/ato2015-2018/2015/lei/l13123.htm.
- Brasil (2016), “Monitoramento dos casos de dengue, febre de chikungunya e febre pelo vírus Zika até a Semana Epidemiológica 03, 2016”, *Boletim Epidemiológico*, Vol. 47 No. 6, pp. 1-10.
- Bruniera-Oliveira, R.B.O., Horta, M.A.P.P., Belo, V.S.S., Carmo, E.H. and Verani, J.F.S. (2014), “Desenvolvimento da Vigilância Epidemiológica de Fronteira no contexto da Globalização: conceitos e marcos teóricos”, *Tempus Actas de Saúde Coletiva*, Vol. 8 No. 3, pp. 75-93.
- Carmo, E.H., Penna, G. and Oliveira, W.K.D. (2008), “Public health emergencies: concept, characterization, preparation and response”, *Estudos Avançados*, Vol. 22 No. 64, pp. 19-32.
- Chaves, M.R.O., Bernardo, A.S., Bernardo, C.D., Dias Filho, J.F., Paula, H.S.C. and Passos, X.S. (2015), “Dengue, Chikungunya e Zika: a nova realidade brasileira Dengue, Chikungunya and Zika: new Brazilian reality”, *NewsLab*, Vol. 132.
- Childs, S. and Mcleod, J. (2004), “Sharing research records and research data: findings from a research project in higher education”, *New Review of Information Networking*, Vol. 10 No. 2, pp. 131-145.
- Cox, A., Verbaan, E. and Sen, B. (2012), “Upskilling liaison librarians for research data management”, *Ariadne*, No. 70, available at: www.ariadne.ac.uk/issue70/cox-et-al.

- Dallmeier-Tiessen, S., Darby, R., Gitmans, K., Lambert, S., Matthews, B., Mele, S., Suhonen, J. and Wilson, M. (2014), "Enabling sharing and reuse of scientific data", *New Review of Information Networking*, Vol. 19 No. 1, pp. 16-43.
- Douglass, K., Allard, S., Tenopir, C., Wu, L. and Frame, M. (2014), "Managing scientific data as public assets: data sharing practices and policies among full-time government employees", *Journal of the Association for Information Science and Technology*, Vol. 65 No. 2, pp. 251-262.
- Dye, C., Bartolomeos, K., Moorthy, V. and Kieny, M.P. (2016), "Data sharing in public health emergencies: a call to researchers", *Bulletin of the World Health Organization*, v. 2.
- Fecher, B., Friesike, S. and Hebing, M. (2015), "What drives academic data sharing?", *PloS One*, Vol. 10 No. 2.
- Gardner, W. (2004), "Compelled disclosure of scientific research data", *The Information Society*, Vol. 20, No. 2, pp. 141-146.
- Garvey, W.D. (1979), *Communication: The Essence of Science: Facilitating Information Exchange among Librarians, Scientists, Engineers and Students*, Elsevier, Burlington, VT.
- Gezelter, D. (2009), "What, exactly, is open science?", *The Open Science Project*, available at: www.openscience.org/blog/?p=269 (accessed October 4, 2016).
- Grant, M.J. and Booth, A. (2009), "A typology of reviews: an analysis of review types and associated methodologies", *Health Information and Libraries Journal*, Vol. 26 No. 2, pp. 91-108.
- Hey, T. and Trefethen, A. (2008), "E-Science, cyberinfrastructure and scholarly communication", in Olson, G., Zimmerman, A. and Bos, N. (Eds), *Scientific Collaboration on the Internet*, MIT Press, London, pp. 15-31.
- Higman, R. and Pinfield, S. (2015), "Research data management and openness: the role of data sharing in developing institutional policies and practices", *Program: Electronic Library and Information Systems*, Vol. 49 No. 4, pp. 364-381.
- Jacomy, M., et al. (2014), "ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi Software", *Plos One*, Vol. 9 No. 6, p. 98679.
- Knight, G. (2015), "Building a research data management service for the London school of hygiene & tropical medicine", *Program*, Vol. 49 No. 4, pp. 424-439.
- Losoff, B. (2009), "Electronic scientific data & literature aggregation: a review for librarians", *Issues in Science and Technology Librarianship*, Vol. 59.
- Lopez, F. (2015), "Workshop de datos científicos", Introducción. In: Conicet, available at: <http://pt.slideshare.net/fernandoariellopez/workshop-de-datos-cientificos-introduccion> (accessed December 7, 2015).
- Melero, R. and Hernández-San-Miguel, J. (2014), "Acceso abierto a los datos de investigación, una vía hacia la colaboración científica", *Revista Española de Documentación Científica*, Vol. 37 No. 4, p. e066.

Mlakar, J., Korva, M., Tul, N., Popović, M., Poljšak-Prijatelj, M., Mraz, J., Kolenc, M., Resman Rus, K., Vesnaver Vipotnik, T., Fabjan Vodusek, V. and Vizjak, A. (2016), “Zika virus associated with microcephaly”, *New England Journal of Medicine*, Vol. 374 No. 10, pp. 951-958.

Musso, D., Roche, C., Robin, E., Nhan, T., Teissier, A. and Cao-Lormeau, V.M. (2015), “Potential sexual transmission of zika virus”, *Emerging Infectious Diseases*, Vol. 21 No. 2, pp. 359-361.

National Institutes of Health (2003), NIH Data Sharing Policy and Implementation Guidance, National Institutes of Health, Bethesda, May 3, available at: http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#goals.

Ruttenberg, S. and Rishbeth, H. (1994), “Sir Granville Beynon: a celebration of his eightieth birthday on 24 May 1994 World Data Centres – past, present and future”, *Journal of Atmospheric and Terrestrial Physics*, Vol. 56 No. 7, pp. 865-870.

Sankoh, O. and Ijsselmuiden, C. (2011), “Sharing research data to improve public health: a perspective from the global south”, *The Lancet*, Vol. 378 No. 9789, pp. 401-402.

Santilli, J. (2015), “Biodiversidade e Conhecimentos Tradicionais Associados: o Novo Regime Jurídico de Proteção”, *Revista do Ministério Público do Distrito Federal e Territórios*, Vol. 9, pp. 21-73.

Søndergaard, T.F., Andersen, J. and Hjørland, B. (2003), “Documents and the communication of scientific and scholarly information: revising and updating the UNISIST model”, *Journal of Documentation*, Vol. 59 No. 3, pp. 278-320.

Strauss, A. and Corbin, J. (2008), “*Pesquisa qualitativa: técnicas e procedimentos para o desenvolvimento de teoria fundamentada*”, Porto Alegre, Artmed.

Sturges, P., Bamkin, M., Anders, J.H., Hubbard, B., Hussain, A. and Heeley, M. (2015), “Research data sharing: developing a stakeholder-driven model for journal policies”, *Journal of the Association for Information Science and Technology*, Vol. 66, pp. 2445-2455.

Thompson, C.A., Robertson, W.D. and Greenberg, J. (2014), “Where have all the scientific data gone? LIS perspective on the data-at-risk predicament”, *College & Research Libraries*, Vol. 75 No. 6, pp. 13-500.

Vasconcelos, P.F.D.C. (2015), “Doença pelo vírus Zika: um novo problema emergente nas Américas?”, *Revista Pan-Amazônica de Saúde*, Vol. 6 No. 2, pp. 9-10.

Walport, M. and Brest, P. (2011), “Sharing research data to improve public health”, *The Lancet*, Vol. 377 No. 9765, pp. 537-539.

Whitmire, E. (2002), “Disciplinary differences and undergraduates’ information-seeking behavior”, *Journal of the American Society for Information Science and Technology*, Vol. 53 No. 8, pp. 631-638.

Whitmire, A.L., Boock, M. and Sutton, S.C. (2015), “Variability in academic research data management practices: implications for data services development from a faculty survey”, *Program*, Vol. 49 No. 4, pp. 382-407.

WHO (2009), “Dengue: guidelines for diagnosis, treatment, prevention and control”, World Health Organization.

WHO (2016), “Latest Zika situation report”, WHO, February 19. available at: www.paho.org/hq/index.php?option=com_content&view=article&id=11669&Itemid=41716&lang=en (accessed February 27, 2016).

WHO (2017), “Zika: response funding”, available at: <http://www.who.int/emergencies/zika-virus/response/contribution/en/> (accessed February 2, 2018).

Whyte, A. and Pryor, G. (2011), “Open science in practice: researcher perspectives and participation”, *IJDC*, Vol. 6 No. 1, pp. 199-213.

Zanluca, C., Melo, V.C.A.D., Mosimann, A.L.P., Santos, G.I.V.D., Santos, C.N.D.D. and Luz, K. (2015), “First report of autochthonous transmission of Zika virus in Brazil”, *Memórias do Instituto Oswaldo Cruz*, Vol. 110 No. 4, pp. 569-572.

Zatz, M. (2000), “Projeto genoma humano e ética”, *São Paulo em Perspectiva*, Vol. 14 No. 3, pp. 47-52.

Further Reading

Oms. (2009), Regulamento Sanitário Internacional, Agência Nacional de Vigilância Sanitária, Brasília, available at: www.anvisa.gov.br/hotsite/viajante/Regulamento_Sanitario_Internacional_vers%C3%A3o%20para%20impress%C3%A3o.pdf (accessed January 9, 2017).