

Declaration of input sources in scientific research: should this practice be incorporated to organizational information management?

José Osvaldo De Sordi

Docente-pesquisador do programa de mestrado profissional em Administração da FACCAMP

Manuel Meireles

Docente-pesquisador do programa de mestrado profissional em Administração da FACCAMP

Cláudia Brito Silva Cirani

Docente-pesquisadora do programa de mestrado e doutorado em Administração da UNINOVE

Márcia Carvalho de Azevedo

Docente do Curso de Administração da UNIFESP

This research studies the declaration of input sources for research in scientific communications, more specifically, whether this practice of the academy may be considered a good example to be followed by organizations. Seven hypotheses address two dimensions of input sources: origin (primary or secondary) and nature (data or information). It appears that the declaration of research inputs in the academy is problematic, mostly incomplete or inaccurate. This does not reduce the importance of this practice; it simply indicates that the academy should not be considered a privileged space, with wide dominance and practice excellence. Nevertheless, the information environment of organizations can learn and benefit from the experience of the scientific academy. From the analyses of the research sample, a set of procedures has been developed, which allowed organizational analysts and researchers to elaborate a complete and accurate analysis of the input sources to be declared in organizational or scientific communication.

Keywords: *Primary source; Secondary source; Research source; Scientific communication.*

Declaração de fontes de insumo da pesquisa científica: esta prática deve ser incorporada à gestão da informação organizacional?

Esta pesquisa estuda a declaração de fontes de insumo da pesquisa científica, mais especificamente, se esta prática acadêmica pode ser considerada um bom exemplo a ser seguido pelas organizações. Sete hipóteses abordam duas dimensões associadas a fontes de informação: origem (primária ou secundária) e natureza (dado ou informação). Observou-se que as declarações de fontes da pesquisa acadêmicas são problemáticas, a maioria feita de forma incompleta ou imprecisa. Isto não reduz a importância da prática, simplesmente indica que a academia não deve ser considerada como um espaço privilegiado, com amplo domínio e excelência na prática. No entanto, o ambiente de informação organizacional pode aprender e beneficiar-se da experiência da academia científica. A partir da análise de textos da academia, um conjunto de procedimentos foi desenvolvido, com o propósito de auxiliar pesquisadores e analistas organizacionais a elaborarem análise completa e precisa das fontes de informação declaradas, seja na comunicação científica ou organizacional.

Palavras-chave: *Fonte primária; Fonte secundária; Fonte de pesquisa; Comunicação científica.*

Recebido em 03.12.2012 Aceito em 21.05.2013

1 Introduction

The transition from an industrial society to an information society is characterized by a wide generation and availability of data and information to society (GOREY; DOBAT, 1996). The expansion of the Internet is an example of this tendency: in December 2011, Twenty-nine million new *web servers* were published, which represent a growth of 4.5% per month, similar to that observed in the remaining months of 2011 (NETCRAFT, 2012). The numbers indicate that the number of *web servers* have duplicated at every 16 months. This scenario also

characterizes the expansion of information which is inaccurate, out of context, outdated, inconsistent, wordy, imprecise, and filled with other characteristics linked with poor quality information (DAVENPORT, 1997).

Within organizations, there is much encouragement for the activities of data collection and information generation. The external environment demands more information and organizational transparency. Among the demanding entities, the following are included: regulatory agencies geared to consumer rights, stock exchanges, shareholders, suppliers, customers, and others (DOSHI; DOWELL; TOFFEL, 2011). Internally, data and information are an essential resource for management activities and business competitiveness (SUND, 2003). The literature on information in organizations points to a paradoxical situation, that is, although we are living in a unique period in terms of availability of information, useful and relevant information is often difficult to be found when needed (EDMUNDS; MORRIS, 2000). A significant part of the problem lies on the emphasis of organizations activities for gathering data, rather than on developing activities aimed at improving data and information quality (EPPLER, 2006).

Effective management for quality information requires specific procedures, some of these relying on traditional practices from the scientific academy. The introduction of summaries to organizational information is one of these procedures (EPPLER, 2006). The adoption by organizations of a summary as part of quality information encouraged the development of summary generator software: *summarizers* (ROBB, 2007). Another way to provide quality to organizational information from the current practices of the scientific academy is through the adoption of stricter methods for generating them. Literature directed to the organizational public is pragmatic in nature, often informal, characterized by an abundance of practical solutions, however weakened by lack of evidence (VAN AKEN; ROMME, 2009). The need for greater attention to the methods used for generation of organizational information is highlighted by Sund (2003, p. 501): "the demand for evidence-based information means that the analysis must be conducted according to the principles of scientific research". For this reason, there are technical and scientific methods, characterized by accuracy and precision, which are applied to the process of generation of organizational information. The *Evidence-based management* approach is an example of practices aimed at satisfying this concern (PFEFFER; SUTTON, 2006).

Another technique used by the scientific academy, which demonstrates care toward information resources, is the process of analysis and declaration of inputs used to generate information. A broad and correct declaration of research inputs involves two dimensions: 1) nature; whether it is data or information, and 2) origin; that is, if it comes from a Primary or Secondary input. In general, preference is given to primary sources, as they are closer and more connected to the object from which information is intended to be generated, which results in better quality information. To assist in the analysis of published scientific

information, it is the practice of the scientific community to declare the inputs used to generate the information used. This practice benefits the scientific community in the numbers of ways: 1) when activities are geared toward research and selection of texts, the researcher can consider the sources declared as criteria for assessing information quality; 2) when the researcher is generating new information, it becomes important for consideration and reflection about the inputs to be used in the process, as well as the quality of information being generated.

The object of this research is the declaration of research inputs by the scientific community. Our objective is to develop a critical analysis of the research inputs declared in scientific articles, in order to investigate if the academic practice is a good example to be applied in business organizations, in the same way as the introduction of summaries and the use of most rigorous methods for analysis and generation of information. Thus, the specific objectives of this research are: 1) analyze the use and accuracy of declaration of the nature and origin research input; 2) propose a template for the analysis of research inputs.

This research is needed not only because organizations are known to collect information of questionable quality (DAVENPORT, 1997), but also because of the influences of the new information society on the behavior of academic and scientific research. There is a growing amount of information and data available on the web sites of Internet which provide data and information from several other entities: web sites, databases, reports and software. This leads to increased complexity and greater effort needed to properly declare the sources available on the Internet, both for those who publish, and for those who use and reference (MICHALOWSKI; THAKKAR; KNOBLOCK, 2005). Although this problem is found in all areas of knowledge, the social science area is the most sensitive, as it often uses mixed methods and multiple resources for research and data collection (PRATT, 2009). Scientific research methods require that researchers analyze the object of research, whenever possible, by crossing multiple sources of data and information (DENZIN, 1989; YIN, 2003; JICK, 1979). This practice is generically referred to as "data triangulation" and is justified in considering that the analysis developed from multiple angles or sources, adds quality to the research, making it more complete, impartial and accurate.

Society's concern with the declaration of sources of the generated information is present in new communication protocols. The Dublin Core Metadata Initiative (2010) is one of the efforts organized to provide interoperability of data between creators and users. DCMI is an international standard for describing information resources, and it is considered an important part of the Internet infrastructure. It consists of a set of 15 elements of metadata, which may be considered the least common denominator for describing information resources (WEIBEL, 1997). Among these 15 metadata, the "source" attribute is observed for describing and recording the inputs used in the research.

2 Concepts and research hypotheses

This section discusses the concept of input declaration in scientific research, divided into four terms derived with two dimensions: primary source versus secondary source (associated with origin dimension); and data versus information (associated with the nature dimension). These four concepts are used in the formulation and logic structuring of the hypotheses of this research.

In the context of scientific research "source" indicates something mentioned in the text that is related therewith, and especially supports the information contained therein (MERRIAM-WEBSTER ONLINE DICTIONARY, 2012). The relation of the source to the object of the research is fundamental to understanding the origin dimension, whether it is primary or secondary. Campos and Cury (1997) use the relational aspect to explain the primary source: primary sources not only provide explanations about the object of research, but they also receive them, and in this case, sources and objects in the system of investigation and research maintain a relation of interdependence. There is a flow of explanations emanating from sources, shifting meanings of objects, and explanations emanating from the object, changing or keeping the meanings presented by the sources. In short, the primary source has a strong relation to the object of research; it is capable of providing explanations about them. The analysis of proximity or distance of the source from the object of the research will indicate the classification of the source as to its origin, whether primary or secondary. According to Solomon, Wilson and Taylor (2007), primary sources of information are those that are closer to the event, the period of time, the individual or other entity which is the object of research.

The distinction between primary and secondary source is subjective and **contextual** (DALTON; CHARNIGO, 2004). The distinction is by no means simple. Considering that the source is just one source in a specific historical context, this same source can be either primary or secondary, depending on what it is used for (KRAGH, 1989). Such information is important, because, in order for a source to be classified, insight is necessary, not only for the object of research, but also for the objective of research. Two examples found during our exploration and analysis of the research articles sample are presented below to illustrate this concept.

Hall's (2002) research aimed to explore how the President of the United States of America, George Bush, used public opinion polls in the process of rhetorical invention in his presidential speeches. Although Hall has not interviewed the president, he did interview important professionals in his team, who were responsible for preparing the President's speeches, and such sources are correctly classified as primary sources of research. These very same professionals in President Bush's team could be classified as secondary sources in a different context, for

example, if the objective of research were to analyze the stress of being president, the source would be secondary. In the first scenario, the respondents are fully involved, that is to say they are responsible for elaborating presidential speech. In the second scenario, however, they can only express their opinion through their perception of president's behavior, since they have not lived the experience of the presidency.

Another example is a study conducted by Turner (2006), who applied osteology (part of anatomy that deals with bones) and phylogeny (genetic succession of organic species) in order to specify a still unknown species of prehistoric crocodile. Among the analyzed sources of evidence, there were four partial skulls of the animal. For the purpose of palate definition (top of the oral cavity) and the internal skull structure, the partial skulls were correctly classified as a secondary source of research. They do not provide direct attributes of these animal's parts; however, they serve as parameters for assumptions about them. If the issue under consider were the skull of the animal, such evidence could be classified as a primary source.

Therefore, it should be noted that the definition of the origin of a source, whether primary or secondary, also depends upon the context. It is a relational issue, which can be described by the following question: how close is the evidence provided by the source to the object of research, according to the objective of such research? It is not about veracity, but proximity, considering that the persons themselves who are objects of the research, could provide incorrect information to the researcher.

Primary sources are preferred because they are directly connected to the object of study. This reduces the risks of using inputs of questionable quality due to discrepancies, errors and noise caused by third parties. Inputs from secondary sources may contain quality problems due to errors, generated at the time of analysis, observation, description, translation or any other form of interaction of the second actor with the primary subject of the research. Thus, for the secondary sources used in a research, there is greater attention in terms of arguments, analyses and methodological procedures to ensure quality of research inputs.

In the academy, there is a perception of higher quality and credibility of primary sources compared to secondary sources (SWINEHART; MCLEOD, 1960). Therefore it is assumed that researchers prefer to declare input sources in their research as being primary. The idea is to convey the perception of better quality of the primary source to the research as a whole, i.e., improving the research value. Therefore, it is expected that articles have a greater incidence of research inputs incorrectly declared as primary sources than research inputs reported incorrectly as secondary sources. This reasoning led to the first research hypothesis:

h_{a1} Research inputs declared as of secondary origin are more accurate regarding the use of the term origin (primary or secondary) than the research inputs declared as of primary origin.

Another important dimension for the classification of research sources is whether the nature of input is data or information. Data are collections of relevant evidence about a fact or entity observed, whereas information is an interpretation of a data set, according to a consensual and relevant purpose for the target audience (BOISOT; CANALS, 2004). Data are abundant and easy to understand, characterized as an input to generate information. As the description of the term indicates, data is something given or accepted as a basis for reasoning or inference (MERRIAM-WEBSTER ONLINE DICTIONARY, 2012). It requires no interpretation and analysis, as it is the case with information, and presents itself as an attribute which is easy to understand and record. On the other hand, information needs to be interpreted and placed correctly in context. This may require analysis and additional efforts for its generation or interpretation. As an example, the interviews conducted by Hall (2002) with team members of President George Bush brought information to the research context: opinions and perceptions of the respondents regarding the context of the questioned object. In article of Turner (2006), the four partial skulls of animals show exact measurements, easily measurable, exemplify research inputs categorized as data.

As with the secondary sources in relation to primary sources, the collection of input characterized as information, compared to data, requires greater care by the researcher. It should be noted that the information was generated from someone's interpretation within a specific context. Thus, the use of input information should be preceded by arguments and care in terms of methodological procedures. A broad declaration of research inputs becomes necessary, stating not only the origin, whether primary or secondary, but also its nature: if it is data or information.

The evolution of Information and Communication Technology resources (ICT) in terms of speed and significance has facilitated the acquisition of data by researchers. The research of Chen, Francis and Miller (2002) research is an example which dealt with primary data from totally inhospitable locations: water temperature readings in various parts of the Arctic Ocean, captured by buoys with sensors which perform data transmission by satellite. The increasing availability and facilities for obtaining data, in particular by advances in ICT, led to the development of the second hypothesis:

h_{b1} Research inputs declared as data are more often used in research than those declared as information.

Historically, the traditional sciences ("*hard sciences*"), such as physics, chemistry, and mathematics work intensively with data collection and the paradigm of positivist research. A greater acceptance of other

research paradigms by the academy, such as constructivist, participant/reclaimable and pragmatic are more recent, having arisen mainly from the seventh decade of the twentieth century (CRESWELL, 2003). Consequently, the dominant practice and terminology in the literature regarding methods of scientific research are still dominated by the positivist culture of data collection. Therefore the use of the terms survey data, collect data and analyze data are very common. Thus, by force of habit, many information inputs end up being declared by researchers as data. This assumption helped to provoke the third hypothesis:

h_{c1} Research inputs declared as data are less accurate in the use of the term which refers to nature (data or information) than the inputs of research declared as information

The perception of preference for the primary input and data motivated the development of a new hypothesis, combined with a category of inputs definitions with greater probability of errors:

h_{d1} The category "primary source of data" is the least accurate of all categories surveyed, both with respect to origin and nature.

Still associated with the prevalence of the positivist culture, in which one works predominantly with input data, we have considered the idea that many declarations of inputs do not mention nature, focusing only on the input origin. Thus the following hypothesis was defined:

h_{e1} The percentage of research that do not point out the nature of the source (whether data or information) is greater than the percentage of research that states both origin and nature.

Complementing the logic of the aforesaid hypotheses, supported by the tradition and the limitations of the dominant positivist culture, a supplementary scenario is devised: researchers who declare the inputs of research more broadly, stating both the dimension nature and the dimension origin of the input are likely to have a more systemic view of the entities involved with the declaration of research inputs, and probably a greater mastery and insight for employing the terms associated with these two dimensions. As the origin dimension is common to all nine categories of articles with sources reported and considered in the analyzes of this study, we used the analysis of this dimension to check if the number of dimensions used in the declaration of source is associated with increased accuracy and mastery in the declaration of research inputs. This logic is shown in the following hypothesis:

h_{f1} Articles that declare not only the origin (primary or secondary), but also the nature of what was collected (data or information), are more accurate in the declaration of the research input origin.

All articles of the research sample were obtained from the PROQUEST and EBSCO electronic databases of scientific articles. In these

databases there is a "relevance indicator" for each article, calculated from the number of citations that the article receives from all other articles present in the same database of scientific articles. This indicator makes it possible to verify if this supposedly greater importance of the article represents greater success in the use of terms associated with the inputs nature and origin dimensions. This observation led to the last hypothesis:

h_{g1} Articles sorted by the databases as being of greater importance (more quoted) are more accurate in the use of the terms associated with dimensions nature (data or information) and origin (primary or secondary) than those of less relevance.

3 Research method

To test our hypotheses, the articles were analyzed and grouped into nine categories of research interest, as shown in the first column of Figure 1. The selection of the articles took place from September 2011 to January 2012. For every search conducted, the following settings in the search software were selected: "*double blind review*" in order to retrieve only scientific articles, and "*full text*" to retrieve only articles which had their full text available. Figure 1 describes the keywords and logical operators used in the researches of the relevant articles, belonging to the nine categories of interest in this study.

Initially, 20 articles for each category of interest were collected, in which 10 were from PROQUEST and 10 from EBSCO databases. During the research process, the lists generated by the search software were sorted by relevance, i.e., the number of citations the article received from other articles from the database. Using the lists of articles sorted by relevance, a random selection of five most and five least relevant was made. Electronic versions of articles selected by chance were copied (*file download*) to one of the nine specific subdirectories of the nine categories of articles described in Figure 1. The descriptive attributes of the articles were recorded in a control spreadsheet, containing the article title, an indicator of greater or lesser relevance, an indicator of origin database and the terms of interest to be analyzed.

The control spreadsheet of the nine categories of interest and the files containing the texts of articles were sent to researchers who, individually, played the role of auditors in relation to the use of terms associated with the dimensions origin and nature of research inputs. It should be stressed that all articles were analyzed by two reviewers. Whenever there was a discrepancy between their opinions, the analysis coordinator forwarded such article to a third reviewer in order to resolve the conflict. The reviewers were selected based on attributes associated with training, experience and productivity in research, such as: having a Ph.D. degree, providing significant scientific production, and involvement in research activities for over ten years.

Figure 1 list the categories of articles with respect to the number of terms associated with the dimensions of research inputs, ranging from one to four terms. For the nine classes of articles, the reviewers initially indicated whether the term (keyword) associated with the origin dimension found in the text, was within the context of interest, i.e., associated with the description of the input source for the research. The term "primary source" for example, presented a very broad and diverse application and therefore, of little relevance to the context of interest for this research. The use of this term as a way to declare the research input source, occurred in only 7% of the analyzed texts, as outlined in Table 1. Among many other semantic uses, the term was used to declare: primary energy source, primary funding source, primary contamination source, among others. For these situations, the article was marked by the reviewer as "out of context".

Fig. 1. Criteria for selection and analysis of the nine categories of articles

Category of articles (Articles which declared inputs using the following search terms)	Criteria for selection of articles (Terms analyzed and logical operators used in the selection criteria of two electronic databases of scientific articles)	Dimensions for the analysis of inputs			
		Origin		Nature	
		Primary	Secondary	Data	Information
Primary data source (PD)	("primary source of data" OR "primary data source") AND_NOT ("secondary source of data" OR "secondary data source")	one term		one term	
Secondary data source (SD)	("secondary source of data" OR "secondary data source") AND_NOT ("primary source of data" OR "primary data source")		one term	one term	
Primary data source & Secondary data source (PDS)	("primary source of data" AND "secondary source of data") OR ("primary data source" AND "secondary data source")	one term	one term	TWO terms	
Primary information source (PI)	("primary source of information" OR "primary information source") AND_NOT ("secondary source of information" OR "secondary information source")	one term			one term
Secondary information source (SI)	("secondary source of information" OR "secondary information source") AND_NOT ("primary source of information" OR "primary information source")		one term		one term
Primary information source & Secondary information source (PISI)	("primary source of information" AND "secondary source of information") OR ("primary information source" AND "secondary information source")	one term	one term		TWO terms
Primary source (PS)	"primary source" AND_NOT "secondary source"	one term			
Secondary source (SS)	"secondary source" AND_NOT "primary source"		one term		
Primary source & Secondary source (PSS)	"primary source" AND "secondary source"	one term	one term		

Source: Authors.

With the term associated with the origin dimension within the context of interest - input source for research - the reviewer analyzed if the definition "primary source" or "secondary source" was correctly

attributed, thus indicating as "term correctly used by the authors" or "term improperly used by the authors". As previously described in the second section, the analysis of the relation between research input and research object has a subjective and contextual nature (DALTON; CHARNIGO, 2004; CAMPOS; CURY, 1997). Thus, this activity required not only the identification of the object, but also the objective of research for each scientific article analyzed, in order to enable an analysis of relation, of distance or closeness, between input source and research object. The core question for this analysis was: can the declared input be considered directly associated with the research object?

Unlike the origin dimension of the research input, whose analysis allowed three response options, such as: "out of context", "term correctly used by the authors", or "term improperly used by the authors", the nature dimension of the input allowed only two answers: "nature correctly defined" and "nature incorrectly defined". The analysis focused on checking whether the entity declared as data was in fact data, and whether the entity declared as information was indeed information. For this analysis, several actions by the reviewers were required, depending on the declaration of source input in the article, such as: a) accessing the Internet Address (URL) indicated for describing the input, b) examining the declaration of structure used to present the input; for example, some used the word 'form' (for data), others used 'report' (for information), c) treatment given by the researchers to the input, for example, content analysis (for information), tabulation (for data). In short, the reviewers developed several actions in order to understand more clearly the nature of the declared input, whether characterized as data or information.

Table I - Number of articles analyzed by category, in the initial and final phases

Category of Articles (articles declaring research inputs using the following terms)	Total Articles with the Term (two databases)	Articles Analyzed at	Articles Out of the Context of	Articles in Context for the	Index of Articles Relevance	Total Number of Relevant Articles (two)
Primary data source (PD)	5,385	24	4	20	83%	4,488
Secondary data source (SD)	478	29	9	20	69%	330
Primary data source & Secondary data source (PDSD)	44	22	2	20	91%	40
Primary information source (PI)	4,853	89	69	20	22%	1,091
Secondary information source (SI)	111	34	14	20	59%	65
Primary information source & Secondary information source (PISI)	21	21	17	4	19%	4
Primary source (PS)	88,877	297	277	20	7%	5,985
Secondary source (SS)	9,116	82	62	20	24%	2,223
Primary source & Secondary source (PSSS)	1,688	84	64	20	24%	402
TOTAL	110,573	682	518	164	= sample	14,628

Source: Authors.

The article selection process from the PROQUEST and EBSCO databases, followed by the analysis of the terms associated with the

nature and origin dimensions of the research input sources continued until the nine categories of included 20 articles within the context of interest. The number articles of the second batch and other subsequent batches were calculated individually for each of the nine classes, considering the result of the last batch analyzed in terms of the relevant articles found. For example: for a category of articles with 80% of relevance in the first batch (16 articles in the context), a second batch with 5 articles was drawn randomly. As shown in Table 1, this process resulted in the collection and initial analysis of 682 articles, in which 518 were outside the context of research interest. The total number of articles analyzed in accordance with the dimensions and with the terms of interest of the research was 164, which is the size of the research sample.

Among the nine categories of articles, the "Primary source of information & Secondary source of information" category (PISI) was the only one in which it was not possible to find 20 articles within the context of research interest. As shown in Table 1, there were only 21 articles in the two databases, containing the four terms researched. Among them, only four articles were within the context of research interest.

Before the reviewers proceeded the analysis of the terms associated with the origin and nature of the research inputs, they went through a training program which addressed the main characteristics to be observed in the declaration of research inputs. To prepare the reviewers training material, it was observed that the literature concerning the designation of primary or secondary origin for the generating source of the research input is focused primarily on the rules from the type of procedure used to collect such input. Few authors addressed the issue of relational analysis between the object of research and the entity that provided the input, and these few do not fail to prescribe procedures for collecting such input. This observation led to an extension of the initial objective of this research, which now includes the creation of a guide (*template*) to analyze the research input with the purpose of developing a comprehensive and correct declaration of it, which constitutes the second specific objective of this research.

4 Data analysis and hypotheses testing

In this section, the tests performed to analyze the hypotheses of this research will be introduced.

h_{a1} Research inputs declared as of secondary origin are more accurate regarding the use of the term origin (primary or secondary) than the research inputs declared as of primary origin.

To test this hypothesis, among the 164 sample articles, we compared 104 articles that declared the term primary source (articles of categories *PD*, *PDSD*, *PI*, *PISI*, *PS* and *PSSS*) with 104 articles that reported the term secondary sources (articles of categories *SD*, *PDSD*, *SI*, *PISI*, *PS* and *PSSS*). The accuracy of secondary sources (86.54%) is

significantly greater than the accuracy of primary sources (59.62%). To ensure the results quality, we used the binomial proportion test, whose result showed that there is a significant difference at the significance level of 0.05 with a p -value equal to 0.000. Consequently, **hypothesis h_{a1} is not rejected**. This result is consistent with the initial expectation which encouraged the development of the following hypothesis: there is a higher incidence of false primary sources than false secondary sources.

h_{b1} Research inputs declared as data are more often used in research than those declared as information.

In the two databases surveyed, EBSCO and PROQUEST, 110,573 articles have been identified which fall into one of the nine categories of articles of research interest. Upon analyzing the relevance of the terms "primary" and "secondary" for the origin dimension, which was an aspect considered in the initial phase of the analysis, an index of article relevance was created for each of the nine categories. The use of this index for the total number of articles initially identified reduced the total number of articles relevant to the terms of the research interest to 14,628, as described in the last column of Table 1.

The relevant articles are classified into nine categories: three of them are exclusively for data nature (PD, SD, PDSD), three others are for information nature (PI, SI, PISI) and there are three that only declare origin (PS, SS, PSSS). Thus, when comparing the relevant articles of the three categories of data nature with the three categories of information nature, the greater use of data nature becomes obvious. As shown in Table 2, the number of articles with data nature (4,858) is greater than the number of articles with information nature (1,160) by a rate of 319%. Therefore, considering the articles that express the input nature, **hypothesis h_{b1} is not rejected**.

Table II - Comparison of the data nature sources with information nature sources in order to test hypothesis h_{b1}

Data Nature Sources		Information Nature Sources	
PD	4,488	PI	1,091
SD	330	SI	65
PDSD	40	PISI	4
Total	4,858	Total	1,160

Source: Authors.

h_{c1} Research inputs declared as data are less accurate in the use of the term which refers to nature (data or information) than the inputs of research declared as information.

To test this hypothesis, we considered 80 terms used in 60 articles of the three categories which declare the nature of the input as data (PD, SD, PDSD) and, in addition to that, the 48 terms present in 44 articles which declared the input nature as information (PI, SI, PISI). The number

of terms in conjunction with the accuracy of the two groups, according to the nature declared is shown in Table 3. Articles, in which the data nature was declared, have shown an accuracy of 60% and others, an accuracy of 81.25%. The binomial proportion test was applied and the result showed that the difference is significant: it can be stated that, at the significance level of 0.05, articles which declared the input nature as information are significantly more accurate (81.25%) than those that declared the input nature as data (60%). Consequently, **hypothesis h_{c1} is not rejected.**

Table III - Comparison of the accuracy between articles that have declared the source nature as data and articles that have declared the source nature as information, for testing hypothesis h_{c1}

Articles which declare the source nature as Data				Articles which declare the source nature as Information			
Category	Number of terms evaluated	Accuracy of the Origin term		Category	Number of terms evaluated	Accuracy of the Origin term	
		quantity	%			quantity	%
PD	20	7	35.00%	PI	20	15	75.00%
SD	20	19	95.00%	SI	20	18	90.00%
PDS	40	22	55.00%	PISI	8	6	75.00%
TOTAL	80	48	60.00%	TOTAL	48	39	81.25%

Source: Authors.

h_{d1} The category "primary source of data" is the least accurate of all categories surveyed, both with respect to origin and nature.

To test this hypothesis, the accuracy of the term "primary source" of the 20 valid articles in the "primary source of data" (PD) was compared to valid articles of other categories that have used the term "primary source": PDS (20 articles), PI (20 articles), PISI (4 articles), PS (20 articles) and PSSS (20 articles). The procedure was the same for the term "Data nature". Table 4 shows the quantities and accuracy of two terms of PD category, comparing them with the same terms used in other categories.

Per Table 4, the accuracy of the category "primary data source" (39.28%) is significantly lower than the accuracy of the other categories (70.97%). The binomial proportion test was used, so the result showed that there is a significant difference at the significance level of 0.05 with a p -value equal to 0.000. Consequently, **hypothesis h_{d1} is not rejected.**

Table IV - Comparison of category PD terms with similar terms under other categories for testing hypothesis h_{d1}

CATEGORY PRIMARY SOURCE OF DATA (PD)			OTHER CATEGORIES		
Term "primary source"					
Category	Quantity	Accuracy	Category	Quantity	Accuracy
PD	20	07	PDSD	20	09
PD	20	07	PI	20	15
PD	20	07	PISI	04	03
PD	20	07	PS	20	14
PD	20	07	PSSS	20	14
Total	100	35	Total	84	55
Term "data nature"					
Category	Quantity	Accuracy	Category	Quantity	Accuracy
PD	20	10	SD	20	19
PD	20	10	PDSD	20	14
Total	40	20	Total	40	33
TOTAL	140	55	TOTAL	124	88
% accuracy		39.28%	% accuracy		70.97%

Source: Authors.

h_{e1} The percentage of research that do not point out the nature of the source (whether data or information) is greater than the percentage of research that states both origin and nature.

As outlined in Table 1 above, 14,628 articles were considered in this research. Of them, 8,610 do not declare the input nature (categories PS, SS, PSSS), i.e., only 58.86% declare only the source origin. The articles that declare the input nature and origin are a total of 6,018 (categories PD, PI, SD, SI, PDSD, and PISI), i.e. 41.14% indicate origin and nature. These data are described in Table 5. Therefore, according to the binomial test, **hypothesis h_{e1} is not rejected**: the percentage of researches that don't indicate the source nature (whether it is data or information) is significantly higher, at the significance level of 0.05, than the percentage of researches that declare both origin and nature.

Table V - Comparison of articles that declare origin and nature with articles declaring only source for testing hypothesis h_{e1}

Articles that have declared input origin and nature		Articles that have declared only the input origin	
PD	4,488	OS	5,985
PI	1,091		
SD	330	SS	2,223
SI	65		
PDSD	40	PSSS	402
PISI	4		
TOTAL	6,018	TOTAL	8,610

Source: Authors.

h_{f1} Articles that declare not only the origin (primary or secondary), but also the nature of what was collected (data or information), are more accurate in the declaration of the research input origin.

The test of this hypothesis involved analyzing 104 articles that, in addition to the origin (primary or secondary), also declare the nature of the input (data or information) (categories PD, PI, SD, SI, PDSD, and PISI). These were compared with the 60 articles declaring only the origin (categories PS, SS and PSSS).

The accuracy in the use of the term origin (primary or secondary) for articles that declare both origin and nature was 67.97%, i.e. significantly lower than the accuracy of the articles that reported only origin: 81.25%. These data are described in Table 6. We used the binomial proportion test whose result showed that there is a significant difference at the significance level of 0.05 with a p -value equal to 0.0356. Therefore, an inverse relation to what was expected was found. Consequently, **hypothesis h_{f1} is rejected.**

Table VI - Comparison of articles that declare input origin and nature with articles that only declare origin for testing the hypothesis h_{f1}

Articles that declare input origin and nature				Articles that declare only input origin			
Category	Number of analyzed terms	Accuracy of the Origin term		Category	Number of analyzed terms	Accuracy of the Origin term	
		quantity	%			quantity	%
PD	20	7	35.00%	PS	20	14	70.00%
PI	20	15	75.00%	SS	20	19	95.00%
SD	20	19	95.00%	PSSS	40	32	80.00%
SI	20	18	90.00%	TOTAL	80	65	81.25%
PDSD	40	22	55.00%				
PISI	8	6	75.00%				
TOTAL	128	87	67.97%				

Source: Authors.

h_{g1} Articles sorted by the databases as being of greater importance (more quoted) are more accurate in the use of the terms associated with dimensions nature (data or information) and origin (primary or secondary) than those of less relevance.

Table 7 shows the accuracy of each of the four key terms (data, information, primary and secondary), according to two groupings of articles indicated by the databases of scientific articles: the most relevant and the least relevant.

Table VII - Shows the accuracy of each of the four key terms (data, information, primary and secondary), according to two groupings of articles indicated by the databases of scientific articles: the most relevant and the least relevant

Term Analyzed	Articles with Lowest Relevance		Articles with Highest Relevance	
	Accuracy in the term use	Non-accuracy	Accuracy in the term use	Non-accuracy
Data	35	8	20	17
Information	16	7	16	1
Primary	36	27	23	14
Secondary	54	8	33	5
TOTAL	141	50	92	37
	191		129	

Source: Authors.

The binomial proportion test was applied to the 320 terms analyzed under the two categories, in which 129 were associated with most relevant articles and 191 with less relevant articles. The results, at the significance level of 0.05, showed no significant difference with a p -value equal to 0.6214 between the proportions of accuracy, both for relevant and irrelevant articles. Among the 129 most relevant articles of different categories, 92 (71.32%) were accurate in the use of the terms nature and origin. This ratio does not differ significantly from the proportion observed with respect to the 191 articles of less relevance of the various categories: 141 (73.82%). Thus, **hypothesis h_{g1} is rejected.**

4 conclusions

Most of the time, the declaration of the inputs for scientific research is partial, as evidenced by the test of hypothesis h_{e1} ; 58.86% of the researches do not report the nature of the research input. In addition to omission of information in regards to the declaration of inputs, there is considerable inconsistency in what is reported. The test of hypothesis h_{a1} indicated that the term "primary source" is incorrectly used 40.38% of the time when it is used to describe a research input. The category "primary source of data", which represents 27% of all articles with declaration of research inputs, is the least accurate of all categories surveyed, as evidenced by the test of hypothesis h_{d1} , since its accuracy is only 39.28%. Thus, it can be stated that the declaration of the research input in scientific articles is incomplete or inaccurate in most cases.

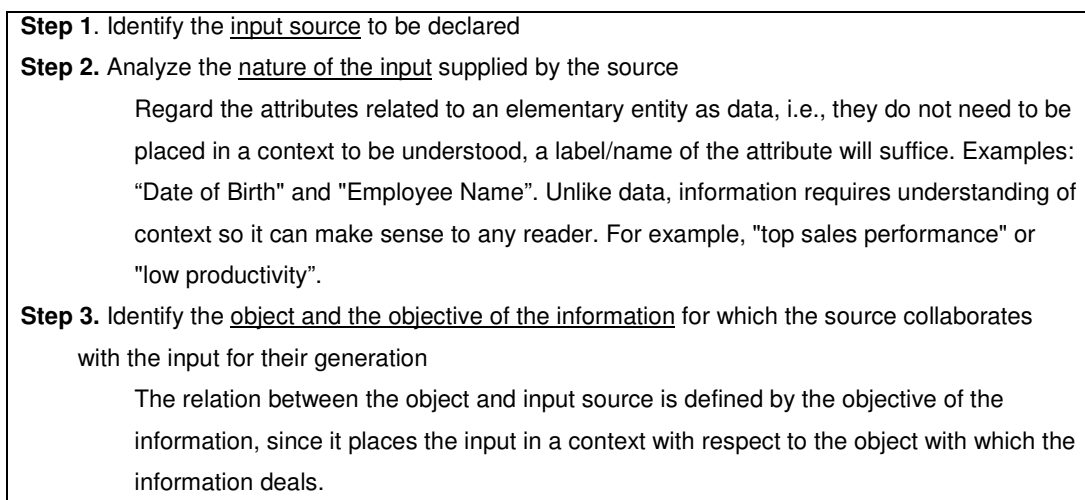
The isolated analysis of the nature dimension indicates that although the information term is less often employed (h_{b1}) it is more accurate (h_{c1}): out of the sources that use the term "data", 40% are incorrect, whereas out of the total number of declarations that use the term information, 18.75% are misused. This misconception can be attributed to the positivist culture of books and materials about research methodology, with strong emphasis on data collection, considering that there is no other

reason that may lead a researcher to intentionally declare data rather than information.

The separate analysis of the origin dimension exhibited a very pronounced difference, both in greater frequency of use and greater misuse of the term "primary source" in regard to "secondary source". In an attempt to better understand this discrepancy, it was observed that research methods books that conceptualize the declaration of research sources, as well as electronic environments for testing the use of these terms, are prescriptive in linking the term primary or secondary to research procedures. For example, if a given input is obtained from interviews, it is defined as a primary source; if documents are collected in the field, they are defined as a secondary source. A few authors address the need for a relational analysis between the object of research and research input to be declared, however, they still prescribe and associate the terms (primary and secondary) to specific procedures for collection or identification of input, rendering these texts incoherent. Another aspect that cannot be dismissed as a partial justification for the observed discrepancy is related to the preference of researchers for the declaration and use of the term "primary source". The idea is that common sense, which is predominant among researchers, assigns greater value and quality to research that works with primary sources. This results in a greater use of the term "primary source", but this aspect has not been extensively analyzed by this research.

This information is relevant to supports our first research specific objective: analysis of declaration use and accuracy of nature and origin research input dimensions in scientific articles. From reviewers' reflections regarding analysis of research inputs declarations, conducted in 164 articles of the sample in this study, a set of recommendations was derived, which is shown in Figure 2. This information speaks to our second research specific objective: to propose procedures to structure the main analyses to be performed by the researcher or by the organizational analyst, in order to declare openly and accurately the inputs used for generating the information.

Fig. 2. Procedures for analysis of origin and nature dimensions of the input



Step 4. Analyze the input origin supplied by the source

Answer the following question: How close is the evidence provided by the source to the object of information, according to its objective? An input should be described as a primary when its generating source is as close as possible to the event, the individual, artifact, process or any other entity understood as object or subject of the information. Otherwise, it should be declared as a secondary source.

Source: Authors.

The rejection of hypothesis h_{g1} is also significant in that it suggests that the articles classified on the basis of scientific communication as being of greatest relevance are not significantly more accurate in the declaration of inputs (nature and origin) than the articles of minor importance. This indicates that the declaration of research inputs does not impact the decision to cite a paper. This is a problem considering that the declaration of undue inputs can affect the quality of scientific research, which in turn can lead to an increased risk of building new knowledge from low quality sources inputs. This is one more aspect which justifies and points to the need for greater commitment from researchers vis a vis the declaration of their research inputs.

In regard to the overall objective of research, i.e. to verify whether the declaration of research inputs by the scientific community can be characterized as a good example to be applied in the context of organizations, two conclusions emerged: a) the declaration of the research inputs in scientific articles is troublesome, incomplete or inaccurate in most cases, and b) there is need for greater commitment from researchers for declaring research inputs. This does not reduce the importance of the practice of declaration of inputs for the information environment, be it in a scientific or in an organizational context. However it demonstrates that the scientific academy should not be considered as a privileged space, with broad dominance and excellence in practice. Although the academy may not be considered a *benchmark* to be widely followed in relation to the practice in question, organizations can certainly learn from the academy. The observation of the concepts involved, the reason for these terms and deploying them through techniques, identifying both correct the improper issues will bring real gains to the information environment of organizations, both for the generation process, and the selection of information.

4.1 Suggestion for continuing of research

The efforts of researchers to obtain research inputs can be compared to the efforts of human evolution studied by anthropologists: nomadic humans only **collected** what nature offered to them, whereas humans, who settled down on land, **harvested** what they planted (DIAMOND, 1999). There are situations in which the researcher simply collects data and information already available. Other information, however, requires more effort for the generation and gathering of inputs

for research, e.g., developing and applying questionnaires, conducting interviews and planning sessions or group work. Similarly, in this study, which addressed the importance of the correct declaration of the dimensions of the research input origin and nature, one can analyze the adequacy of treatment given by the researchers to the gathered inputs and collected inputs. The analysis applies in terms of relevance and coherence to two instances: actions associated with the process of obtaining input and actions associated with the preparation of such input for its effective use. One of the expected results is to identify a set of more relevant and recommended activities for the collected inputs, as well as another set of inputs to the gathered inputs, in accordance with research techniques and methods used.

References

- BOISOT, M.; CANALS, A. Data, information and knowledge: have we got it right? *Journal of Evolutionary Economics*, v. 14, n. 1, p. 43-67, 2004.
- CAMPOS, E. N.; CURY, M. Z. F. Fontes primárias: saberes em movimento. *Revista da Faculdade de Educação*, v. 23, n. 1, p. 1-7, 1997.
- CHEN, Y.; FRANCIS, J. A.; MILLER, J. R. Surface temperature of the arctic: Comparison of TOVS satellite retrievals with surface observations. *Journal of Climate*, v. 15, n. 24, p. 3698-3708, 2002.
- CRESWELL, J. W. *Research design: Qualitative, quantitative, and mixed methods approaches*. 2. ed. Thousand Oaks, CA: Sage Publications, 2003.
- DALTON, M. S.; CHARNIGO, L. Historians and their Information sources. *College & Research Libraries*, v. 65, n. 5, p. 400-425, 2004.
- DAVENPORT, T. *Information ecology: mastering the information and knowledge environment*. New York, NY: Oxford University Press, 1997.
- DENZIN, N.K. *The research act: a theoretical introduction to sociological methods*. 3. ed. New York, NY: Prentice Hall, 1989.
- DIAMOND, J. *Guns germs and steel: the fates of human societies*. New York, NY: WW Norton & Company, 1999.
- DOSHI, A. R.; DOWELL, G. W. S.; TOFFEL, M. W. *How Firms respond to mandatory information disclosure* (Working Paper No. 12-001). 2011. Disponível em: <<http://ssrn.com/abstract=1879248>>. Acesso em: 22 nov. 2011.
- DUBLIN CORE METADATA INITIATIVE. *Dublin Core Metadata Element Set. Version 1.1*. 2010. Disponível em: <<http://dublincore.org/documents/dces/>>. Acesso em: 17 mar. 2011.
- EDMUNDS, A.; MORRIS, A. The problem of information overload in business organisations: a review of the literature. *International Journal of Information Management*, v. 20, n. 1, p. 17-28, 2000.

EPPLER, M. J. *Managing information quality: increasing the value of information in knowledge-intensive products and processes*. 2. ed. New York, NY: Springer, 2006.

GOREY, R.M.; DOBAT, D.R. Managing in the Knowledge Era. *The Systems Thinker*, v.7, n.8, p.1-5, 1996.

HALL, W. C. "Reflections of yesterday": George H. W. Bush's instrumental use of public opinion research in presidential discourse. *Presidential Studies Quarterly*, v. 32, n. 3, p. 531-558, 2002.

JICK, T. D. Mixing qualitative and quantitative methods: triangulation in action. *Administrative Science Quarterly*, v. 24, n. 4, p. 602-611, 1979.

KRAGH, H. *An introduction to the historiography of science*. Cambridge, MA: Cambridge University Press, 1989.

MERRIAM-WEBSTER ONLINE DICTIONARY. *Source*. 2012. Disponível em: <<http://www.merriam-webster.com/thesaurus/source>>. Acesso em: 2 dez. 2011.

MICHALOWSKI, M.; THAKKAR S.; KNOBLOCK, C. A. Automatically utilizing secondary sources to align information across sources. *AI Magazine*, v. 26, n. 1, p. 33-44, 2005.

NETCRAFT. *Web Server Survey*. December 2011. Disponível em: <<http://news.netcraft.com/archives/category/web-server-survey/>>. Acesso em: 10 dez. 2011.

PFEFFER, J.; SUTTON, R.I. Evidence-based management. *Harvard Business Review*, v. 84, n. 1, p. 62-74, 2006.

PRATT, M.G.. For the lack of a boilerplate: tips on writing up (and reviewing) qualitative research. *Academy of Management Journal*, v.52, n.5, p.856-862, 2009.

ROBB, D. How search is converging with business intelligence. *Business Communications Review*, v. 37, n. 8, p. 28-31, 2007.

SOLOMON, A.; WILSON, G.; TAYLOR, T. *100% information literacy success*. Clifton Park, NY: Thomson Delmar Learning, 2007.

SUND, R. Utilisation of administrative registers using scientific knowledge discovery. *Intelligent Data Analysis*, v. 7, n. 6, p. 501-519, 2003.

SWINEHART, J. W.; MCLEOD, J. M. News about science: channels, audiences, and effects. *Public Opinion Quarterly*, v. 24, n. 4, p. 583-589, 1960.

TURNER, A. H. Osteology and phylogeny of a new species of *Araripesuchus* (Crocodyliformes: Mesoeucrocodylia) from the Late Cretaceous of Madagascar. *Historical Biology*, v. 18, n. 3, p. 255-369, 2006.

VAN AKEN, J. E.; ROMME, G. Reinventing the future: adding design science to the repertoire of organization and management studies. *Organization Management Journal*, v.6, n.1, p.5-12, 2009.

WEIBEL, S. The Dublin Core: a simple content description model for electronic resources. *Bulletin of the American Society for Information Science*, v. 24, n. 1, p. 9-11, 1997.

YIN, R. K. *Case study research: design and methods*. Thousands Oaks, CA: Sage Publications, 2003.