Looking for the identity of Information Science in the age of big data, computing clouds and social networks [1]

Mauricio, Barcellos Almeida

Federal University of Minas Gerais, mba@eci.ufmg.br
Av. Antônio Carlos, 6627
Belo Horizonte, Brazil

Renato, Rocha Souza

Getúlio Vargas Foundation, renato.souza@fgv.br
Praia de Botagofo, 190
Rio de Janeiro, Brazil

Renata, Baracho Porto

Federal University of Minas Gerais, renatabaracho@eci.ufmg.br
Av. Antônio Carlos, 6627
Belo Horizonte, Brazil

Abstract

In this paper we discuss, under a critical point of view, the current Information Science landscape and some future prospects regarding contemporary information phenomena. We present thoughts about the process of thematic deflation of Information Science, through the analysis of the research objects currently under development in this field. In addition to this, we look at the process of absorption of these and other relevant objects in distinguished knowledge fields. We seek to challenge the emphasis and the volume of interdisciplinary research within the field, and present some comments about what might be the results of such processes for the future of

[1] This work is partially supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Governo do Estado de Minas Gerais, Brazil, Rua Raul Pompéia, nº101 - São Pedro, Belo Horizonte, MG, 30.330-080, Brazil.

Information Science. Subsequently, we analyze the impact in the Information Science field due to phenomena like information boom, the consolidation of the social networks as interactive spaces, cloud computing, as well as other key elements.

Keywords: Information Science. Epistemology. Interdisciplinary Studies. Information Technology.

Introduction

Within all scientific fields, and even among those more recently established, one can say that Information Science (IS) is one of the most introspective fields with regards to research themes. The conceptual issues that underlie IS come up with very passionate arguments that their real research objects are sometimes blinded or faded into the background. Considering that seminal texts of the IS field are so broad – as seen in the works by Bush (1945) or Shanon and Weaver (1949) – and departing from such multifaceted concepts – again, as seen in Barlow (1994), Hofkirchner (1999) and Capurro (2003) – it is surprising that some authors are still able to find certain agreement about what IS should encompass under its umbrella (Borko, 1968; Zins, 2009). Indeed, the consolidation process of a relatively recent research field like IS, one find many obstacles if compared to the same processes in the so-called "hard-sciences".

With respect to this relatively unique characteristic in the IS field, Wersig (1993) emphasizes the need for a pragmatic-evolutionary perspective along with attempts at theoretical construction. This should be done so that both empirical research objects and interfaces with other fields, work as a support system for the permanent construction and re-construction of the whole research field. Wersig (1993) describes the IS professional through an analogy with the "weaving bird". It is a kind of bird that builds its nest by weaving elaborate and complex nets, which refers to the thematic connections, required in the ordinary work of IS professionals.

As IS is a dynamic research field, working on research objects somehow fluid and ubiquitous, we present here an opinion about how the paradigmatic and epistemological transformation occurs. There is a need for this field when considering the scenario of new technologies and new ways to handle

information. We openly assume our intention of producing a certain provocation, as long as we believe that there is an on-going deflation of IS field as an autonomous research field. This deflation has occurred because of the increasing overlapping among research objects of IS and other fields. Such phenomenon has resulted in a substantial reduction of interest in the scope of IS, as well as the migration to other fields, whose research objects could be considered legitimate IS research objects.

The remaining part of the article is organized according to three scenarios: the deflation, the transformation and future perspectives.

Deflation

Borko (1968, p.3), in its remarkable definition that attempts to capture IS epistemology and praxis, asserts:

> Information Science is that discipline that investigates the properties and behavior of information, the forces governing the flow of information, and the means of processing information for optimum accessibility and usability. It is concerned with that body of knowledge relating to the origination, collection, organization, storage, retrieval, interpretation, transmission, transformation and utilization of information. This includes the investigation of information representations in both natural and artificial systems, the use of codes for efficient message transmission, and the study of information processing devices and techniques such as computers and their programming systems. It is an interdisciplinary science derived from and related to such fields as mathematics, logic, linguistics, psychology, computer technology, operations research, the graphic arts, communications, library science, management, and other similar fields.

Considering this definition, one might claim that virtually all research objects of the contemporary academic research would involve at least one component connected to IS. However, one can identify a bias, albeit broad, which can be read between the lines. The main concerns of IS are processes

of information and knowledge representation, in addition to the manipulation of the records produced along these processes. The nature of these processes presupposes that information suffers successive abstractions, modeling and representations so that it can be organized, transmitted, codified and consumed. Buckland (1991) captures these processes and related objects by considering all of them as manifestations of the concept of information. Furthermore, it is worth emphasizing that information and knowledge representation are seminal subjects in the scope of IS.

Despite some discussion, there is a reasonable agreement about both the moment and the reasons for the birth of IS. On one hand, IS arises to deal with the issues that appeared from the increasing complexity in the management of informational collections. On the other hand, it fulfills the need for the creation and adaptation of methodologies – the praxis. Both aforementioned facts come from the fields of documentation and librarianship. However, the last few years of the 20th century were characterized by the fast expansion of information technology and by the acceleration of the processes of information transformation. These processes include the creation, representation, storage, organization, dissemination and consumption of information. This meant the issues from decades ago originally justifying the appearance of a so-called "Information Science", are nowadays exponentially felt.

Networks, mobile devices, tablets and other gadgets, as well as digital libraries and emergent developments like wearable computing, have continuously changed contexts, shorten cycles, and reinvented material supports. These events have redefined the relation between people and information records. There is a progressive "undocking" process, chagig deeply the usual connections between information and its supports for recording, exchanging and consuming. Organization of the large mass of data needs new and creative solutions, suggesting the real need of a science to deal with information, namely IS.

This "technological ecology" could favor a new inspiration to the IS field, by multiplying the informational issues and the consequent demand for new solutions. But what has really occurred is the gradual migration of genuine IS research objects to other fields. To illustrate such thematic migration, one can mention socio-technical artifacts for information retrieval (like digital

libraries), instruments for knowledge representation (like ontologies), techniques for domain modeling, information systems studies in their several contexts of use, among others. All of these examples of research objects have been actively studied by other sciences in fields of Linguistics, Administration, Computer Science and Information Systems.

On the other hand, themes like "social responsibility", "inter and transdisciplinarity", "information and work", "information society", and so forth (ANCIB, 2009), although relevant and legitimate themes for a social and applied science, have directing the field to focus on aspects related to the social issues coming from Sociology, Anthropology, History, among others. The rise of research in themes related to management, as Knowledge Management and Competitive Intelligence make IS close to fields of Administration and of some subfields of Engineering. All these examples show borderline overlapping that contribute to dissolving IS frontiers and to the lack of its identity.

With regard to the processes of information and knowledge representation, which are strongly connected to the origins of the IS field, one can expect that it become a mere and passive spectator in a highly dynamic environment. Even though IS is a dominant field when the subject is the study and construction of instruments for information organization, like indexing languages and information systems interfaces, these themes are nowadays associated with "technical sciences". Indeed, these sciences have incorporated these instruments into their constructs. Borko (1968) analyses the research objects mentioned in the "Current Research and Development in Scientific Documentation" and lists nine categories, namely:

1. Information need and uses
2. Document creation and copying
3. Language analysis
4. Translation
5. Abstracting, classification, coding and indexing
6. System design
7. Analysis and evaluation
8. Pattern recognition
9. Adaptive systems

Among these objects, which are very current topics, only 1, 2 and part of 5 continue to be recognized as part of the core of IS. Linguistic analysis has been properly developed in fields like Computational Linguistics, Corpus Linguistics, Computer Science and even Applied Mathematics. Topics 3 and 4, as well as part of 5 and 7, have been developed under the label of Natural Language Processing. Computer Science and Engineering encompass topic 6, namely, Systems Design. Finally, the growing field of Artificial Intelligence has developed the topics 8 and 9.

In many universities around the world, the "schools of information science" have been transformed into "schools for information studies", where special groups gather the required profiles able to exert the socio-technical interdisciplinary skills. In the United States, an example of this sort of discussion is the so-called *i-school´s movement*, in which several renowned universities have adopted new techniques, attitudes and even new curricula in schools and departments (Detlefsen, 2008). In addition, the conferences of the International Society for Knowledge Organization (ISKO 2014) also denote a clear direction that seeks a broad coverage, while maintaining the theoretical focus. The incorporation of themes that would suggest a social bias, as the *folksonomies*, occur in the scope of the knowledge organization systems, a very dear and traditional research subject within IS.

We conclude this first scenario emphasizing the need to take care of IS field, so that it does not become a mere niche among other fields. Some hope that information professionals, instead of being the Wersig´s weaving bird, become a sort of "remora"[2], which feeds on the thematic leftovers of topics that other fields develop.

Transformation

There are surveys and research available in the literature, which characterize IS under countless different aspects, for example Capurro (1991) and Whittaker (2011), to mention a few. The scenarios presented here in this section do not correspond to an exhaustive literature review. Instead, we

[2] Parasite fishes that have on the head a sucking disk with which they attach themselves to sharks and other fishes so that they can take rests of food.

chose just to approach the effects of some new developments in IS, namely, Big Data, Cloud Computing and Social Networks. We believe that those developments may have caused impacts and paradigmatic changes in IS research.

Social Networks

Social networks are structures that gather actors – individuals and institutions – through links (Easley et al, 2010), which can be reified as social or technological arrangements. The representation and study of those networks include, but are not limited to, fields of Anthropology, Biology, Communication, Economy, Geography, Information Science, Psychology, Sociology and Socio-Linguistics.

Here, we only approach the social networks constructed from socio-technical arrangements, as is the case of some available on the web, for example, networks for relationships like Twitter[3] and Facebook[4]. In its early days, Twitter was discarded because it was not considered a relevant information source, but in 2010 its economic messages of 140 characters started to be collected[5] by the United States Library of Congress. In 2014, more than 500 million messages were exchanged in Twitter each day[6]. Similarly, in 2014 Facebook had around 1.1 billion users exchanging an amazing volume of information[7]. Only these two instruments completely surpassed the information currently disseminated in printed or digital media. The main feature of this sort of information is the fluidity, dynamics and thematic ephemerality, which are explored in activities of environmental monitoring, sentiment analysis (Liu, 2012) and even the recording of human history.

It is worth mentioning another phenomenon, Wikipedia, one of the most remarkable crowd sourced collective developments of our age. Wikipedia is a free and collaborative multilingual encyclopedia, which makes it possible

[3] https://twitter.com/

[4] https://pt-br.facebook.com/

[5] http://www.businessinsider.com/library-of-congress-is-archiving-all-of-americas-tweets-2013-1

[6] http://www.internetlivestats.com/twitter-statistics/

[7] http://www.statisticbrain.com/facebook-statistics/

to copy, change and extend any type of information. It is distinguished from blogs because it does not use the traditional concept of authorship. Anyone with Internet access is able to modify any Wikipedia article, and each reader is a potential collaborator. In addition, Wikipedia goes beyond the results obtained by search engines, as the response to a query is a single text.

It is still worth remembering that social networks and their related phenomena point out to the ephemeral character of both the informational records and the instruments required to organize these records. Indeed, these collective constructions allow one to create products of unimaginable extent, which will require new approaches and studies (Surowiecki, 2004).

Big data

The explosion of information, which is sometimes called information or data deluge, is more than a large increase in the number of records. The term "big data" has become somewhat commonplace: it is used to refer to any mass of data, which cannot be processed without specific computational structures. In addition to the large volume, some definitions assigned to big data features like variety, velocity, and veracity (Laney, 2012). Indeed, since the 1990s large volumes of data have been manipulated with the aim of producing simulations and predictions. However, the textual non-structured information has been increasing a lot since the 2000s. This phenomenon has had a strong impact on the academic community and in research, most notably in fields like IS. The professionals using traditional techniques of subject analysis, classification and categorization have been trying to adapt their expertise, both to the gigantic collections, and to the speed of production of summaries and indexes (Magnusson; Vanharanta, 2003; Nodus Labs, 2012).

In the scope of academic communities, works of literature review, which are essential for research, are increasingly complex and in some cases impractical. Often a researcher has to cut arbitrary and explicit excerpts from the literature, meaning that the quantity of publications and available sources about specific subjects has become intractable. In other scientific cases, it is now a common research mode called "data driven research" (Moe-Behrens, 2012). According to this mode, one should perform several tests of hypotheses in large masses of data as an exploratory preamble. When promising possibilities are found from the data analysis, it would be possible

to propose qualitative and quantitative enhancements though interactive changes of focus.

Cloud computing

Since we have already briefly described both the social phenomenon materialized through networks, and explosion of information illustrated through the hype around the big data, it is worth mentioning the importance of clouding computing. This phenomenon adds one more abstract level to the process of dematerialization of information records. Some time ago, still without confidence, people changed from physical supports to digital ones like tablets and mobile devices. But now, the cloud computing separates the users from those devices that store their data. Ubiquitous and often freely available storage services (DropBox[8], Google Drive[9], iCloud[10], to mention a few) are part of the lives of millions of people. These people seen unconcerned with the potential lack or theft of both photos and private data, even considering the current quantity of security incidents.

So, one can notice that: (a) social networks have disrupted the geographical limits of connectivity; (b) the new explosion of information has eliminated the possibility for an individualized treatment of information records; and (c) cloud computing has broken up the tangibility of records. So, in this complex context, an important issue begs for a proper answer: what will be the research agenda for IS in the 21st century?

Future Perspectives

What would be the role of IS in a context in which it seems to have lost primacy as a research field, and even centrality, in the thematic of information and knowledge representation? Resuming our initial provocation in the aforementioned scenario of IS deflation, what will be the possible arrangements capable of avoiding such deflation? Revisiting the definition of IS by Borko (1968), as "an interdisciplinary science that investigates the

[8] https://www.dropbox.com/

[9] https://drive.google.com/

[10] https://www.icloud.com/

properties and behavior of information, the forces that govern the flow and use of information and the techniques of processing ", one can conclude that this program has become too broad for the IS field. Indeed, one can even conclude that such a program is too broad for any research field.

Yet, the already cited proposal of Wersig (1993) has become more and more present. It recognizes the new role of knowledge in contemporary society, and advocates the adoption of pragmatism to deal with problems in IS. Thus, in this scenario, epistemological approaches should receive less emphasis. Wersig´s third model for IS suggests, on one hand, a theoretical structure that reduces the attempts of formulating general laws. On the other hand, it suggests the creation of more action strategies from an approach based on the interlacement of scientific concepts. Then, it would weave a proto-net of basic concepts in IS, from which other people or groups could continuously find and integrate new concepts. In this sense, this conceptual net would become more inclusive and strong, in addition to having augmented it´s scientific character.

Examining this from another perspective, on one hand, interdisciplinary studies may weak IS as a research field. But, on the other hand, this same interdisciplinary studies may bring the possibility, and even the prerogative, of mediation among disciplinary dialogues. Such interdisciplinary essence fosters the professional and the researcher of IS to navigate in new theoretical spaces, to adapt to the technological contexts and to continuously reinvent itself.

Curriculum Vitae

Mauricio Barcellos Almeida holds a PhD in Information Science and currently acts as Associate Professor in the Department of Information Theory and Management at the Federal University of *Minas Gerais*, Brazil.

Renato Rocha Souza holds a PhD in Information Science and currently acts as Adjunct Professor in the School of Applied Mathematics at the *Getulio Vargas* Foundation, Brazil.

Renata Baracho Porto holds a PhD in Information Science and currently acts as Adjunct Professor in the Department of Information Theory and Management at the Federal University *of Minas Gerais*, Brazil.

References

ANCIB. (2009). History of the Brazilian Conferences of Research in Information Science. Retrieved July 31, 2011 from http://www.ancib.org.br/enancib/historico-do-enancib/

Borko, H. (1968). Information science: what is this? American Documentation, v.19, p.03-05.

Bush, V. (1945). As we may think, in the atlantic monthly. Retrieved August 22, 2010 from http://ccat.sas.upenn.edu/ jod/texts/vannevar.bush.html

Capurro, R. (2003). The concept of information. Annual Review of Information Science and Technology (ARIST), v.37, n.8, p.343-411.

Capurro, R. (1991). What is Information Science for? a philosophical reflection. In: Vakkari, P.; Cronin, B. (Ed.). Conceptions of Library and Information Science. Proceedings... London: Taylor Graham, p.82-96.

Detlefsen, E. G. What´s is a name? the I-Schools project. Medical Library Association News, v.28, n.3, p.1-17, 2008.

Easley, D.; Kleinberg, J. (2010). Networks, Crowds, and Markets: reasoning about a highly connected world. Cambridge: University Press.

Hofkirchner, W. (1999). The quest for a unified theory of information. In: Hofkirchner, W. (Ed.). Proceedings of the 2nd Intern. Conf. on the Foundations of Information Science. Amsterdam: Gordon & Breach, p. 9-30.

Ingwersen, P.; Järvelin, K.. (2005). The turn: integration of information seeking and retrieval in context. Netherlands: Springer, v. 18.

International Society For Knowledge Organization. ISKO Chapters. Retrieved March 29, 2014 from http://www.isko.org/chapters.html

Laney, D. (2012). The Importance of 'Big Data': a definition. Retrieved March 22, 2014 from https://www.gartner.com/doc/2057415/importance-big-data-definition

Liu, B. (2012). Sentiment Analysis and Opinion Mining, Chicago: Morgan & Claypool.

Magnusson, C.; Vanharanta, H. (2003). Visualizing sequences of texts using collocational networks. In: 3rd Intern. Conf. Leipzig. v.2734, p. 276-283.

Moe-Behrens, G. (2012). From hypothesis to data driven research. Retrieved April 2, 2014 from http://www.sciencetogrok.com/2012/12/from-hypothesis-to-data-driven-research.html

Nodus Labs. (2012). Textexture: visualize any text as a network. Retrieved July 11, 2013 from http://textexture.com/

Osterberg, G. Update on the Twitter archive at the Library of Congress. Retrieved July 3, 2014 from http://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress

Shannon, C.E.; Weaver, W. (1949). The Mathematical theory of communication. Urbana: University of Illinois Press.

Surowiecki, J. (2004). The Wisdom of Crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations. New York: Anchor.

Vakkari, P.; Cronin, B. (1992). Conceptions of Library and Information Science. Historical, empirical and theoretical perspectives. London: Taylor Graham.

Wersig, G. (1993). Information science: the study of postmodern knowledge usage. Information Processing & Management, v.29, n.2, p.229-239.

Whittaker, S. (2011). Personal information management: from information consumption to curation. Annual Review of Information Science and Technology. v.45, n.1, p. 1-62, 2011.

Zins, C. (2009). Knowledge map of information science: issues, principles, implications. Retrieved September 30, 2012 from http://www.success.co.il/is/index.html