UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Marília

OFICINA UNIVERSITÁRIA
CULTURA ACADÊMICA
Editora

Female leadership in the emerging field of digital curation

María-Antonia Ovalle-Perandones
Mirelys Puerta-Díaz
Daniel Martínez-Ávila

Como citar: OVALLE-PERANDONES, María-Antonia; PUERTA-DÍAZ, Mirelys; MARTÍNEZ-ÁVILA, Daniel. Female leadership in the emerging field of digital curation. *In*: JORENTE, Maria José Vicentini; SAN SEGUNDO, Rosa; MONTOYA, José Antonio Frías; MARTÍNEZ-ÁVILA, Daniel; SILVA, Stephanie Cerqueira. (org.). **Digital Curation and Gender in Information Science:** access and preservation. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica; San José: Ediciones Edicic, 2022. p. 37-50. DOI: https://doi.org/10.36311/2022.978-65-5954-328-1.p37-50.

CHAPTER 2

FEMALE LEADERSHIP IN THE EMERGING FIELD OF DIGITAL CURATION*

MARÍA-ANTONIA OVALLE-PERANDONES
Universidad Complutense de Madrid

MIRELYS PUERTA-DÍAZ
Universidade Estadual Paulista

DANIEL MARTÍNEZ-ÁVILA
Universidad de León

ABSTRACT

In this text we analyze the field of digital curation from a gender perspective. We conducted an analysis of the domain until 2020 using mainly bibliometric techniques. We worked with the main reference source in the area, according to the scientific consensus, that is the International Journal of Digital Curation (Miguel et al., 2013), as well as international databases to identify other publications that cited this important source. We obtained, from these sources, an exhaustive list of all the authorship data to identify the gender of the authors (male or female) and the concept of leadership based on the corresponding author.

Keywords: Digital curation, domain analysis; female leadership.

* This chapter is part of the AI-GenBias Project funded by the State Research Agency (Spain) (PDI2019-106695RB-100/ AEI/10.13039/501100011033).

https://doi.org/10.36311/2022.978-65-5954-328-1.p37-50

RESUMEN

En este capítulo se analiza el campo de la curación digital con una perspectiva de género. Se realiza un análisis del dominio hasta 2020 haciendo uso, principalmente, de técnicas bibliométricas. Se trabaja con la principal fuente de referencia en el área según consenso, International Journal of Digital Curation (Miguel et al., 2013), y bases de datos internacionales para identificar otras publicaciones que citan esta importante fuente. De estas fuentes se obtuvo una relación exhaustiva de todos los datos de autoría para después identificar el género (masculino o femenino) y trabajar con el concepto de liderazgo a partir del autor de correspondencia.

Palabras clave: Curación digital, análisis de dominio, liderazgo femenino.

1 INTRODUCTION

Digital curation or content curation is an emerging professional field, emerging in the context of the web content publishing proliferation, and Web 2.0. The Digital Curation Center (DCC) defines the concept as one that has the quality of bringing together the different tasks of management, classification and organization of digital assets throughout their useful life, from the moment of their conceptualization, to their use as active elements, and caring for their preservation so that they can be presented and used in the long term from their preservation site (Rusbridge *et al.*, 2005). In the professional manifesto of content curation (Bhargava, 2009), the object is defined as the search, grouping, organization and sharing of the most relevant content on a specific subject.

As a professional field increasingly linked to research due to the very nature of using electronic environments to disseminate research, DCC provides expert advice and practical help on how to store, manage, protect and share digital research data. It offers a wide range of resources, including online tools, guidance and training. And it provides consulting services in aspects such as policy development and data management planning. As a knowledge domain, digital curation has been disputed by the areas of digital marketing, journalism and communication, engineering and computing, information and documentation or education, among others (Guallar *et al.*, 2020). In any case, since the object of study is linked to technologies, women are at risk of being excluded from their scientific sphere, traditionally rooted in rationalist and positivist positions to which the dominant classes had privileged access.

In this sense, the aim of this chapter is to analyze the scientific leadership of women in the emerging field of digital curation. For this, the main source used is the electronic scientific journal International Journal of Digital Curation (IJDC), specialized in the publication of works, articles and news about the digital object preservation and other related topics. Although some works carried out bibliometric studies of academic production on digital curation (for example, GUALLAR *et al.*, 2020), no previous study has been carried out with a gender perspective.

On International Women's Day, March 8, 2021, the value

of female leadership was highlighted. In the context of any field or domain, scientific leadership measures the production of a given unit of measurement as the main contributor, that is, the part of the production in which the author of correspondence is in that unit of measurement (Moya-anegón, 2012). The present chapter aims to show whether this value transcends in a field as emerging as digital curation or if, on the contrary, the barriers we associate with a particular tradition remain, whether structural or cultural (Meiksins *et al.*, 2019), distancing from a certain scientific equity between the two genres (Palomba, 2006).

2 METHODOLOGICAL PATH

Methodologically, the investigation uses bibliometric techniques to analyze the digital curation domain with special emphasis on women's role, contributions and characteristics in the epistemic community that make up the analyzed domain. Domain analysis as a theoretical-epistemological paradigm was introduced in Information Science in 1995 (Hjørland & Albrechtsen, 1995). Subsequently, Hjørland (2002; 2017) listed bibliometrics as one of the ways to approach domain analysis and has been satisfactorily worked on in several studies (Smiraglia, 2015).

The present work uses bibliometrics to analyze the digital curation domain based on the scientific production of the period 2007-2020. As primary research source it works with the scientific journal specialized in the field of digital curation, the International Journal of Digital Curation. In the context of Information Science and domain analysis, several studies analyzed a domain from the production in a specialized journal (Smiraglia, 2012; Miguel *et al.*, 2013; Guimarães *et al.*, 2015; Oliveira *et al.*, 2017; Alves *et al.*, 2019; Martínez-Ávila; *et al.*, 2020). The relationship between journals and domains is explained by Smiraglia (2015, p. 9) in the following terms:

> Journals are the formal venues for most scholarly communication, and studying them as whole works is also one means of identifying productive elements of a research front. Of course, few journals are devoted to topical areas that are narrowly defined as most domains under study. For example, even in the field of knowledge organization, the principle journal Knowledge

> Organization is devoted to the entire field. Thus it would likely be the most cited journal in all domains within KO, but there are no journals devoted to specific narrow aspects of KO, such as "integrative levels", "multilingual thesauri" or "ethics in KO".

The second main source of the research was Google Citas, used to identify the publications that cite the works published in this specialized magazine. The combination of the two sources provided an exhaustive list of all authorship data and the identification of the genre associated with the authors' names (male or female) to analyze women's scientific leadership according to the role of the corresponding author (Moya-Anegón *et al.*, 2013).

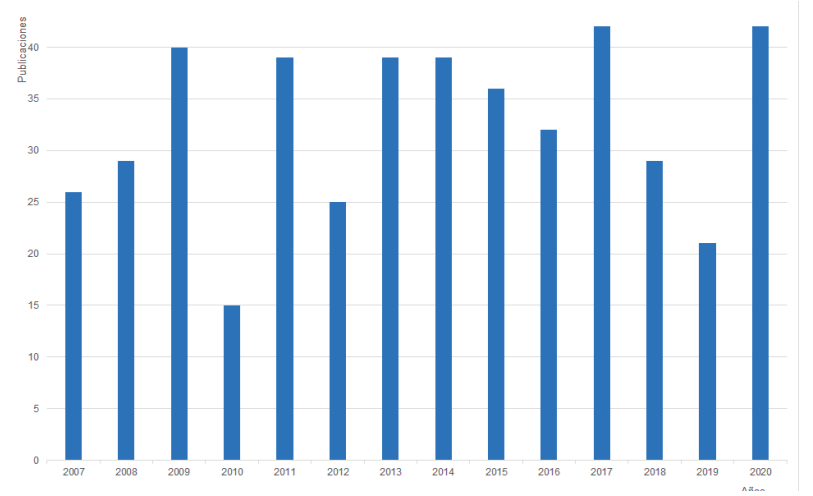
To obtain the records, the free software Publish or Perish (PoP) v.7 was used. PoP allows extracting publications directly from Google Scholar, where the journal is indexed. The used search strategy consisted of placing the title of each publication between quotation marks in the PoP *title words* field. Subsequently, with a bibliographic manager, the data were cleansed and some data omissions were completed by consulting the primary source available on the journal's official website. The authors' gender was identified by consulting the Gender API platform, whose database contains 6,084,389 validated names from 191 different countries. Gender data were exported in .csv format for further processing and integration into the co-authorship network.

3 CHARACTERISTICS OF THE FIELD OF DIGITAL CURATION

3.1 Temporal evolution

In the IJDC journal, 454 works were published in the period from 2007 to 2020. This means that, on average, 32 works are published annually. The trend is to publish a similar number of works, although in some years such as 2010 or 2019 there was a decrease justified by the publication of a single issue in the year. However, a few years ago, even having published a single issue, a large number of publications were concentrated in it, as in the year 2020.

Figure 1: Frequency of publications per year (2007-2020)



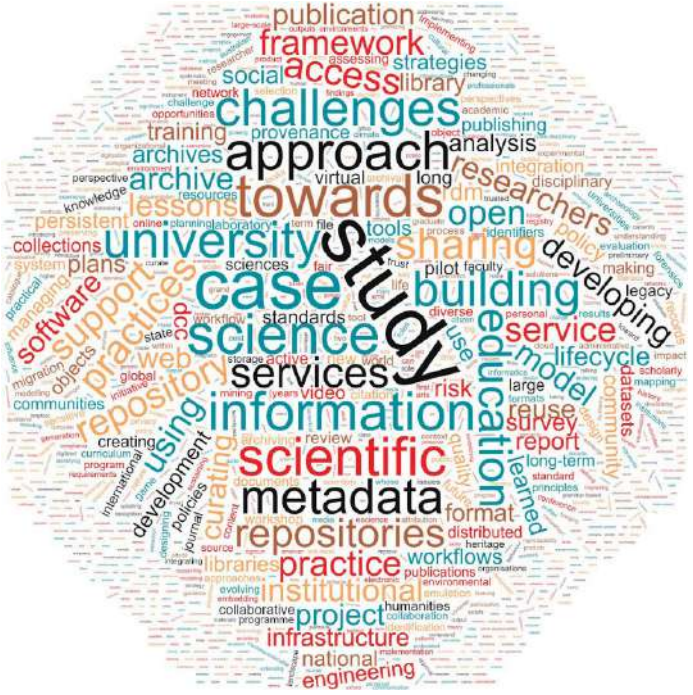
Source: by the authors with Microsoft Excel.

3.2 Title terms

The word cloud of terms used in the titles of all works published in the IJDC is presented, without considering empty words and terms that, although more frequent, are those expected to be found considering the theme of the journal. The most frequent terms, preserving the original language, are: data (273), digital (107), research (98), curation (84), management (73) and preservation (70).

The authors confirm that the thematic coverage of the journal can be defined, both by those terms not included in the word cloud, in addition to *study, case, towards, science, approach, information, scientific, challenges, university, building, metadata* and *education*. Figure 2 includes a total of 1257 title terms. Size meets frequency and is represented with a different color to improve understanding of the cloud.

Figure 2: Cloud of terms included in the titles of articles published in the IJDC



Source: by the authors with Word Cloud Generator.

3.3 Authors' Production

The research elite was calculated based on Price's law, considering that "the number of prolific producers is equal to approximately the square root of the total number of authors in the field" (Price, 1976, p. 3). The formula for √1044 researchers corresponds to approximately the 32 most productive authors. Table 1 (female authors in bold) lists the 21 authors with more than 5 publications.

Table 1: List of authors that make up the research elite

Author	**Total of publications**
Jones, Sarah	11
Ball, Alexander; **Lyon, Liz**; Matthews, Brian	9
Ludäscher, Bertram; Whyte, Angus; Donnelly, Martin;-Treloar, Andrew; **Callaghan, Sarah**	8
Brown, Geoffrey	7
Snow, Kellie; Pryor, Graham; Abrams, Stephen; Carlson, Jake; **Molloy, Laura**	6
Day, Michael; Tedds, Jonathan; Mayernik, Matthew S.; **Willoughby, Cerys;** Knight, Gareth; Missier, Paolo	5

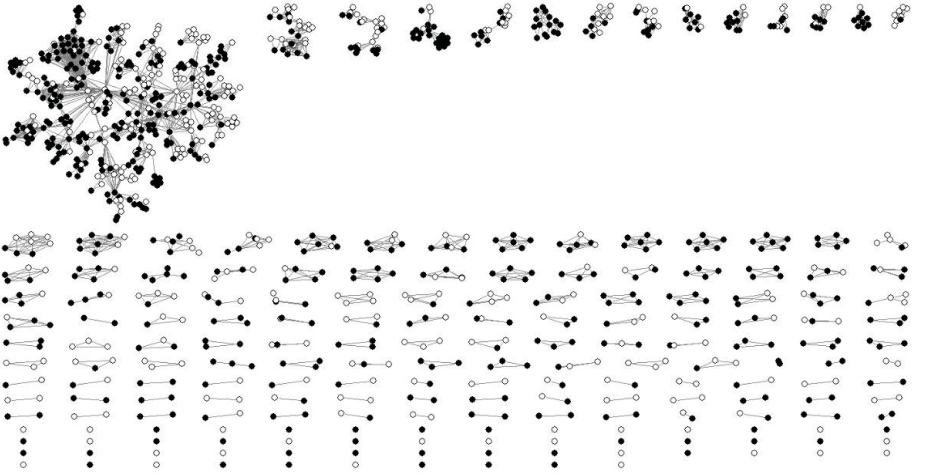
Source: by the authors.

Among the most productive authors in this emerging field, the English researcher Sarah Jones from the Digital Curation Center stands out, with a total of 11 publications. Her studies address the main concerns regarding data management, especially specializing in research on the implementation of data management policies and plans in institutional contexts.

3.4 Co-authorship Networks

The co-authorship network is composed of 1044 nodes. The relationship matrix is symmetrical and weighted. Symmetrical as an author cannot co-author with others without the others being co-authors of that author. It is weighted, considering the number of occasions on which co-authorship between authors occurs. Punctual relationships or with a single document create 2483 relationships in this network; in two or more documents, 182. These authors establish a total of 2665 relationships. From the gender point of view, Figure 3 represents in black the nodes with male authors (628 nodes); and in white the nodes for female authors (416).

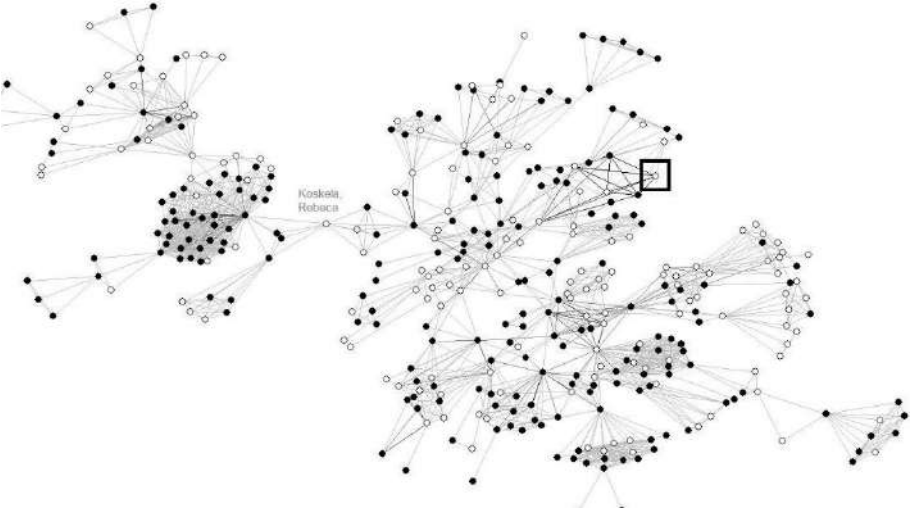
Figure 3: Co-authorship network of the journal IJDC for the period 2007-2020



Source: by the authors with Pajek.

In the same figure and at the bottom, the isolated nodes are shown. In this network, isolated nodes correspond to authors who have never collaborated with any author. This occurred with 52 authors. In the upper left, the main component is clearly identified. Among the 192 components or subnets of the entire IJDC co-authorship network, 380 nodes are connected to this component.

Figure 4: Main component of the IJDC co-authorship network in the period 2007-2020



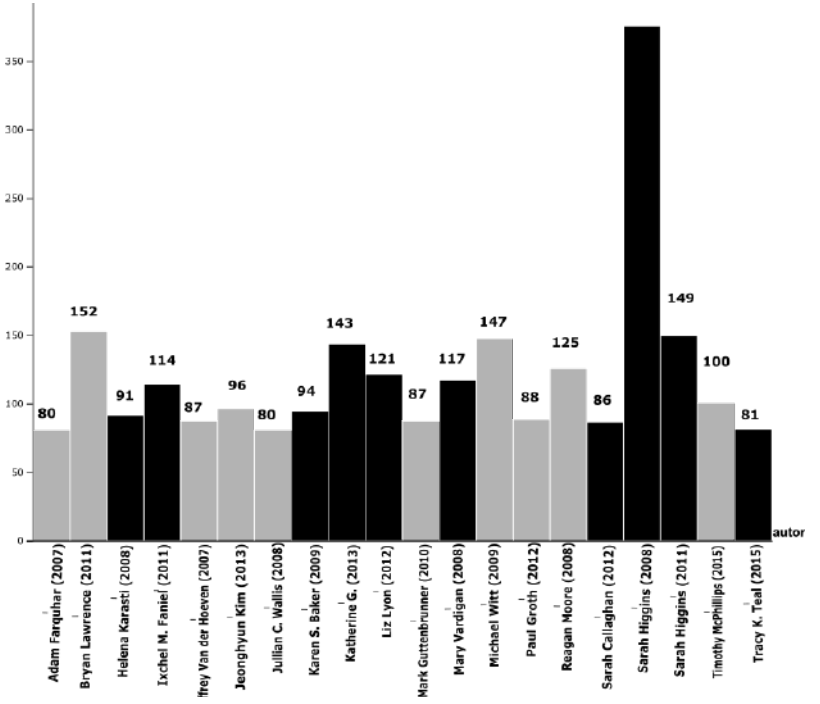
Source: by the authors with Pajek.

As indicated, the subnetwork included in Figure 4 is composed of 380 nodes (226 linked to the male gender; 154 to the female gender) and a weak link is clearly identified in white (corresponding to the author Koskela, Rebeca, once she disappeared, the network would become more fragmented). Rebeca Koskela fulfills the function of connecting the upper left group, which is smaller in size, with the right group, which is larger in size. This author connects with author Bertram Ludascher, the author with the highest degree in the network (49). As was the case across the whole network, the most frequent co-authorship relationships in the network occur on one occasion in 1560 of them (light gray); while on more than one occasion it occurs in 140 relationships (dark gray). Thus, the co-authorship network generated by author Kirsty Merret, in the upper right corner (marked by a small square), is the one that occurs the most.

3.5 Citation dynamics

Figure 5 (female authors in black and male authors in gray) shows the works that received more than 80 citations, identified on the X axis by reference. If we analyze the impact of publications, among the 20 most cited works, 6 articles received more than 100 citations and whose authorship composition is led by the female gender. In this indicator, author Sara Higgins obtains the first and third highest number of citations received for her works. Discussions about the emergence of Digital Curation as an emerging discipline and its proposal for a lifecycle model constitute theoretical pillars of this scientific field.

Figure 5: List of works with more than 80 received citations



Source: by the authors with RawGraph and Inkscape.

3.6 Science and Leadership

The analysis of women's scientific leadership in the emerging field of digital curation as the main contributors to the domain allows us to affirm that women have been authors of correspondence (Moya-Anegón, 2012) in greater proportion than male authors. Analyzing the 92 articles published in the IJDC in the triennium 2018-2020, in 37 of them the corresponding authors were men (third column of Table 2), and 54 were women (second column of Table 2), in addition to an article whose corresponding author was an institution.

Table 2: Scientific leadership in the works published in the IJDC in the 2018-2020 triennium

Year	Scientific leadership women	Scientific leadership men
2020	25	16
2019	10	11
2018	19	10

Source: by the authors.

4 CONCLUSION

Gender disparities in science are a reality (Larivière *et al.*, 2013), however, the digital curation domain shows signs of a small shift in this trend. In the analyzed period, Sarah Jones is the most productive author; Rebeca Koskela plays the role of a node to expand the cohesion of the network; and Kirsty Merrett often works collaboratively given the bond derived from co-authorship; the fact that they are all women contradicts the expected trend (Kwiek & Roszka, 2020). The year 2020 is highlighted by the important change in the dynamics in relation to the IJDC authors as it was the moment when they most frequently performed the functions of scientific leadership.

REFERENCES

Alves, B.H., Dalessandro, R.C., & Santos, F.B.d. (2019). Colaboração Científica no Periódico Knowledge Organization: Elementos para Caracterização de um Domínio. In T.H.B. Barros & N.B. Tognoli (Eds.), *Organização do Conhecimento Responsável: Promovendo Sociedades Democráticas e Inclusivas* (pp. 137-144). Ed. da UFPA.

Bhargava, R. (2009). *Manifesto For The Content Curator:* The Next Big Social Media Job Of The Future?. https://www.rohitbhargava.com/2009/09/manifesto-for-the-content-curator-the-next-big-social-media-job-of-the-future.html

Guallar, J., Codina, L., & Abadal, E. (2020). La investigación sobre curación de contenidos: análisis de la producción académica. *Ibersid,* 14(1), 13-22.

Guimarães, J.A.C., Martínez-Ávila, D., & Alves, B.H. (2015). "Epistemic Communities in Knowledge Organization: An Analysis of Research Trends in the Knowledge Organization Journal." In *ISKO UK biennial conference* 13th–14th July 2015, London.

Hjørland, B. (2002). Domain analysis in information science: Eleven approaches – traditional as well as innovative. *Journal of Documentation,* 58(4), 422-462 https://doi.org/10.1108/00220410210431136

Hjørland, B. (2017). Domain Analysis. *Knowledge Organization,* 44(6), 436-464.

Hjørland, B., Albrechtsen, H. (1995). Toward a new horizon in information science: domain-analysis. *Journal of the American Society for Information Science,* 46(6), 400-425.

Kwiek, M., & Roszka, W. (2020). Gender Disparities in International Research Collaboration: A Study of 25,000 University Professors. *Journal of Economic Surveys*. doi: 10.1111/joes.12395

Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Bibliometrics: Global gender disparities in science. *Nature News,* 504(7479), 211.

Martínez-Ávila, D., Ibekwe, F., & Bochi, F. (2020). The Epistemic Communities and Evolution of Knowledge Domains: A Domain Analysis of the Journal Education for Information. In: M. Lykke, T. Svarre, M. Skov, & D. Martínez-Ávila (Eds.), *Knowledge*

Organization at the Interface (pp. 264-273). Ergon Verlag. https://doi.org/10.5771/9783956507762-264

Meiksins, P., Layne, P., Beddoes, K., Lewis, M., Masters, A.S., & Deters, J. (2019). Women in Engineering: A Review of the 2018 Literature. *SWE Magazine.* https://alltogether.swe.org/2019/04/women-in-engineering-a-review-of-the-2018-literature/

Miguel, S., Hidalgo, M., Stubbs, E., Posadas, P., & Jaureguizar, E.O. (2013). Estudio bibliométrico de género en la paleontología de vertebrados. El caso de la revista argentina Ameghiniana (1957-2011). *Investigación Bibliotecológica,* 27(61), 133-155.

Moya-Anegón, F. (2012). Liderazgo y excelencia de la ciencia española. El *Profesional de la Información,* 21(2), 125-128. http://dx.doi.org/10.3145/epi.2012.mar.01

Moya-Anegón, F., Guerrero-Bote, V.P., Bornmann, L., & Moed, H.F. (2013). The research guarantors of scientific papers and the output counting: a promising new approach. *Scientometrics,* 97, 421-434.

Oliveira, E.F.T.d, Alves, B.H., Prado, M.R.d., & Pavanelli, M.A. (2017). Produção Científica e Inserção Internacional da Revista Scire no Período de 2006 a 2014. *Scire,* 23(1), 47-56.

Palomba, R. (2006). Does gender matter in scientific leadership? *Scientific Careers,* 133-137.

Price DS. (1976). A General Theory of Bibliometric and Other Cumulative Advantage Processes. *Journal of the American Society for Information Science,* 292-306. doi: 10.1002/asi.4630270505

Rusbridge, C., Burnhill, P., Ross, S., Buneman, P., Giaretta, D., Lyon, L., & Atkinson, M. (2005). The Digital Curation Centre: A Vision for Digital Curation. In: 2005 IEEE *International Symposium on Mass Storage Systems and Technology* (pp. 31-41). IEEE. doi: 10.1109/LGDI.2005.1612461

Smiraglia, R.P. (2012). Shifting Intension in Knowledge Organization: An Editorial. *Knowledge Organization,* 39(6), 405-408.

Smiraglia, R.P. (2015). *Domain Analysis for Knowledge Organization:* Tools for Ontology Extraction. Oxford: Chandos Publishing.