Digital Humanities and Open Science: Initial Aspects

Fabiane Führ(✉) and Edgar Bisset Alvarez

Universidade Federal de Santa Catarina, Florianópolis, Brazil
fabiane.fuhr@ufpr.br, edgar.bisset@ufsc.br

Abstract. The digital humanities have acquired more and more visibility and their field of action has expanded due to the increasing digitalization and the large volume of data arising from these processes. Collaborative researches on the impact that is in line with the dimensions of open science impact the scientific production chain. It aims to identify which aspects of open science are approached in the publication regarded to digital humanities. To achieve the general objective it indicates some specific objectives: it identifies the scientific production about open science and digital humanities indexed in the databases Web of Science (WoS), Scopus, and Scientific Electronic Library Online (SciELO) and; it describes how open science is approached in each paper from the corpus and how it is related to digital humanities. It uses the bibliographic manager Zotero to organize the bibliographic data and it uses the software Atlas.ti to the qualitative analysis and applies the data mining tool Sobek and Voyant Tools in the data. From the 13 papers analyzed, only 3 do not use projects or programs related to digital humanities to present the discussion. The data mining tools do no show the relation between digital humanities and open science. It shows the importance of data management and the necessity to have guiding documents. It also points to the relevance of metadata pattern, to work to make the data suitable to FAIR principles, to train researchers and citizens to promote collaboration among different institutions and people that have a diverse background to value open science.

Keywords: Digital humanities · Open science · Data mining · Bibliometric · Scientific production

1 Introduction

Digital Humanities (HD) is a comprehensive field of research, which has been considered by some as a set of practices, by others as a new field of study or a new discipline, or just a new look for research in the Humanities [1].

Digital Humanities gained visibility with the publication of the book *A companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth, released in 2004. However, its origin dates back to research developed in the 1940s and 1950s by Father Roberto A. Busa, who created the Index Thomisticus, which is

E. Bisset Álvarez (Ed.): DIONE 2021, LNICST 378, pp. 154–173, 2021.
https://doi.org/10.1007/978-3-030-77417-2_12

considered the first application of computing to linguistic studies in the works of São Tomás de Aquino [2].

According to Dalbello [3], the HD timeline begins with the introduction of computational methods in the literary, philological, and philosophical fields in the 1950s and continues with a focus on searchable multimedia files and structured texts in the 1980s and 1990s, in addition to editing project texts and electronic collections from the past decades [2000s].

The field of research in Digital Humanities has grown exponentially due to the increasing digitization of documents, which has generated a large mass of data, expanded the number of collaborative research, and changed the scientific production chain. However, Digital Humanities projects are not restricted only to accessibility and dissemination of knowledge. Still, they also concern about the ways of creating and disseminating them [4], the convergences between Digital Humanities and Information Sciences [5–8]. Bibliometric studies on the subject are examples of studies that have been developed [9–11].

This same chain of scientific production has changed due to the movement in favor of Open Science. This movement has been consolidated in many countries since the movement for Open Access, whose decisive landmark was the Budapest Declaration published in 2002. Open Science aims to share and access publications and research data, promote the scientific process's opening, assist in the transfer of knowledge, expand the social and economic impacts of science, and reinforce the social responsibility of science [12].

Thus, Open Science is a movement that proposes new forms of collaborative, interactive, and shared production of information, knowledge, and culture [13]. This movement has gained more and more visibility and has been extended to the most diverse investigation areas.

In this context, it is important to observe in publications on Digital Humanities how the theme of open science has developed. Therefore, this article's general objective is to identify which aspects of Open Science are being addressed in publications referring to Digital Humanities, in the Web of Science (WoS), Scopus, and Scientific Electronic Library Online (SciELO) databases.

The specific objectives are:

a) Identify the scientific production on Open Science and Digital Humanities indexed in the Web of Science, Scopus, and SciELO databases from 2014 to 2020;
b) Describe how Open Science is approached in each of the articles of the research corpus and how they relate to the Digital Humanities.

One of the motivators for the growth of research and publications on Open Science is believed to have been the development of the movement for open access to knowledge, which had its starting point in the Budapest Declaration of 2002. However, all changes require a period for their maturation, which can be seen in this research that aims to analyze when and how open science came to be among the proposals and projects related to Digital Humanities.

2 Methodology

To identify scientific production in the area of Digital Humanities and relate it to Open Science, the expression [("digital humanities" AND "open science")] was defined as a search strategy, but as the databases also have articles published in Portuguese and Spanish, the strategy for the search expression [("humanidades digitais" OR "humanidades digitales" OR "digital humanities") AND ("ciência aberta" OR "ciencia abierta" OR "open science")].

The SciELO Network was opened in 1998, with the creation of the SciELO Brazil Collection. The project was an initiative of the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and the Latin American and Caribbean Center on Health Sciences Information (BIREME/PAHO/WHO). In 2002, the project started to be financially supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [14]. Currently, the SciELO Network has national collections from the following countries: Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Spain, Mexico, Paraguay, Peru, Portugal, Uruguay, South Africa. They are also under development three new collections: Ecuador, Venezuela, West Indies. As it is an open-access program for international cooperation in scientific and academic communication, it is believed that its coverage is representative of the present study.

The Scopus and WoS databases are international databases maintained by the major information companies Elsevier and Clarivate Analytics. Both strategies were applied in the Scopus, Web of Science (WoS), and Scientific Electronic Library Online (SciELO) databases on July 20, 2020. The search was carried out in the title, keyword, and summary fields. There was no application of a filter concerning the publication date, as it is a recent topic. Thus, it would be possible to obtain a greater number of records for analysis.

Both strategies returned the same results since the database's index not only articles in English but also in Spanish and Portuguese. A filter was applied to these results so that only scientific articles were retrieved. Thus, 21 articles were retrieved, 9 articles from the Scopus database and 10 from the Web of Science. The year of publication of these articles corresponds to the years 2014 to 2020. The 2 articles retrieved from the SciELO database were published in 2020.

The bibliographic manager Zotero was used to collect, store, and organize the articles retrieved in the bibliographic survey. After the organization, 8 duplicate files were excluded. Thus, 13 articles remained for analysis.

Of the 13 articles that make up the research corpus, 30.77% are found only in the Scopus database, and another 30.77% only in the WoS. 23.08% are found in Scopus, and WoS databases, and another 15.38% are found in Scopus, WoS, and SciELO databases. Table 1 shows the distribution of articles by database.

Table 1. Number of articles per database

Database	Number of articles
SciELO, Scopus, WoS	2
Scopus, WoS	3
Scopus	4
WoS	4

Source: The authors (2020)

Regarding the publication date, 2019 was the year with the highest number of publications (5), 2017, 2018, and 2020 have 2 publications, and 2014 and 2015 have 1 publication per year.

For qualitative analysis of this research's corpus, the software ATLAS.ti[1] was used, a software for qualitative data analysis, which allows analyzing texts, audios, photos, videos, and working with documents in different file formats [15]. The *Sobek,*[2] text mining tool, was developed from an algorithm initially created by Schenker in 2003 and modified by the Research Group Gtech.Edu at the Federal University of Rio Grande do Sul to make the more accessible tool to educational practice, analyzing relevant words in a text and representing them graphically [16]. Canadians Stéfan Sinclair at McGill University and Geoffrey Rockwell at the University of Alberta developed the *Voyant Tools*[3] app. The application allows you to work with text or text collections and perform basic mining functions. One advantage is that it allows you to work with documents in different formats, supports large volumes of texts, allows the interaction between tools that facilitate navigation and exploration of different scales, among other options useful to the researcher [17].

3 Presentation and Analysis of Results

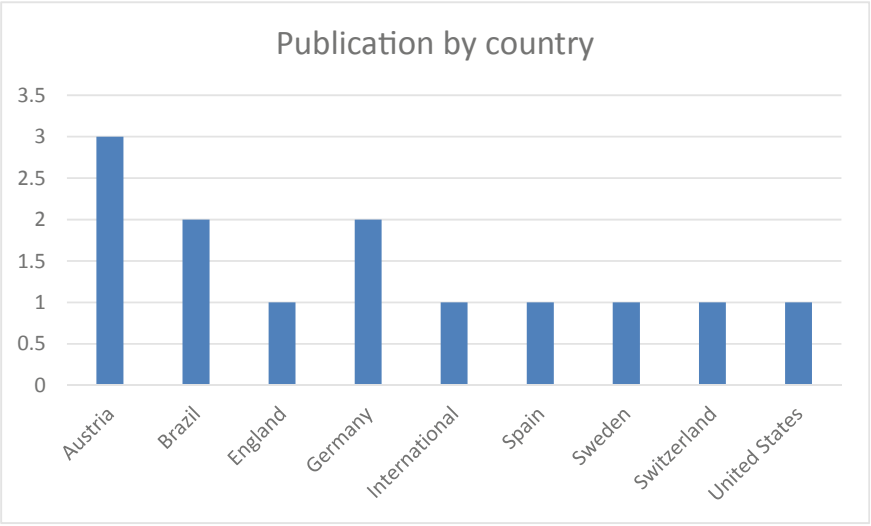
The research was carried out in 13 articles, of which 5 were written in English, and another 5 in German. 2 articles were written in Portuguese and 1 in Spanish. Regarding the country of publication of the journal, it was observed that Austria is the most productive country with 3 publications. Germany and Brazil have 2 publications, and the other countries only 1 publication (Sweden, USA, Spain, Switzerland, England, and one International[4] publication), as can be seen in Graph 1.

[1] https://atlasti.com/.

[2] http://sobek.ufrgs.br/#/.

[3] https://voyant-tools.org/.

[4] ERCIM News magazine is registered with the ISSN as an international publication, as it is a publication of the European Research Consortium for Informatics and Mathematics, which includes research establishments from different European countries [18].



Graph 1. Publication by country

The journals were analyzed concerning Qualis Capes (Quadriênio 2013–2016). And about the impact factor (Journal Impact Factor) provided by the Journal Citation Reports of the Web of Science Group and the journals that were not identified in the Journal Impact Factor, were analyzed to verify if they were on the list of the Emerging Source Citation Report, of Web of Science Group, both Clarivate Analytics [19]. Regarding CiteScore, which is a metric that indicates the impact of the survey. CiteScore is developed by Scopus, Elsevier's database [20].

Table 2. Analysis of qualis, impact factor, emerging source and citescore of journals

Periodicals	Number of articles	Qualis[a]	Journal impact factor	Emerging source citation report	CiteScore
Bibliothek - Forschung und Praxis	1	*	*	Yes	*
Cataloging and classification quarterly	1	A2	*	Yes	1.2
ERCIM news	1	*	*	Yes	*
Estudos históricos	2	B1	*	Yes	0.1
Information research	1	A1	0.763_2019		1.7

(*continued*)

Table 2. (*continued*)

Periodicals	Number of articles	Qualis[a]	Journal impact factor	Emerging source citation report	CiteScore
Literary and linguistic computing	1	A1	1.125_2016		N/A
Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare (VÖB)	3	*	*	*	0.3
Profesional de la Informacion	1	A1	1.580_2019	*	2.1
Publications	1	B5	*	Yes	1.8
Zeitschrift für Germanistik	1	*	*	**	N/A

[a]Qualis from Literary and Linguistic Computing refers to the area of Linguistic and Literature assessment. The others belong to the Communication and Information area.
* There was no information about the journals in the sources consulted
** This journal is indexed in the collection: Art & Humanities Citation Index

It was observed that the German-language journals do not present information in Qualis Capes, which can be justified by the linguistic barrier, since Qualis Capes reflects where teachers and researchers in the area have published the results of their research [21].

To analyze the corpus of this research, the software ATLAS.ti was used, a software for qualitative data analysis, which allows the analysis of texts, graphics, audios, and videos. After reading the articles in advance, 25 categories were established for marking and coding during the reading. The categories were grouped into Families, as assigned by the software.

For the Digital Humanities Family, 16 categories were established (Fig. 1) for marking and analysis. Figure 1 shows 17 nodes (1 representing the Digital Humanities Family and the others representing the 16 categories).

Open science was also the subject of research in the articles. Thus, the Open_Science Family was designated, divided into 9 categories (Fig. 2) for marking and further analysis.

All articles were analyzed for their insertion in the area of Digital Humanities and Open Science. Thus, Table 3 presents the 13 selected articles, their authors, and the projects and programs related to Digital Humanities that are described and addressed in the publication.

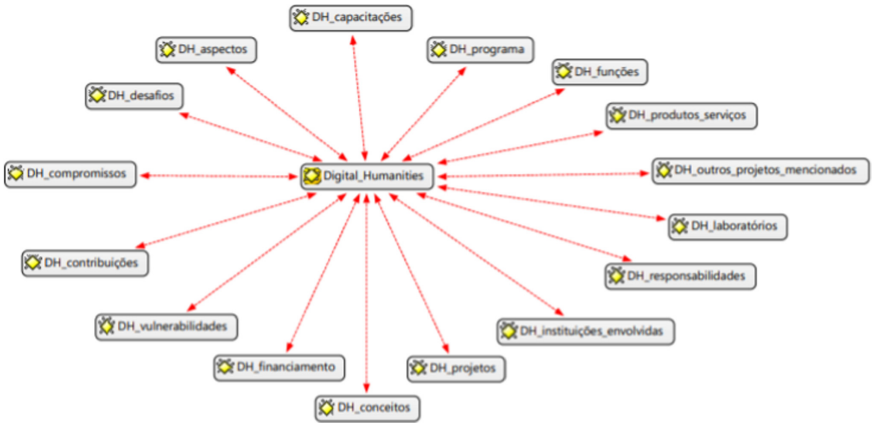


Fig. 1. Digital_Humanities family

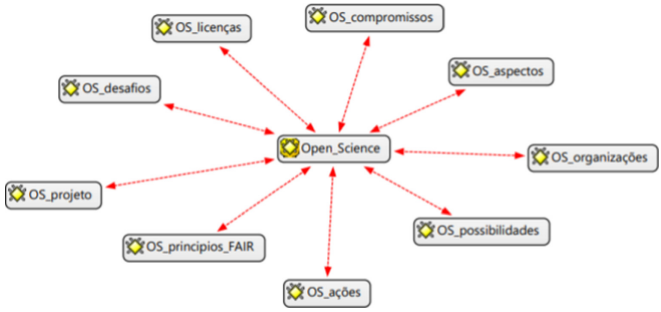


Fig. 2. Open_Science family

Regarding the type of authorship, it was observed that 07 articles are of unique authorship. The other articles have shared authorship distributed as follows: 3 articles have 3 authors and only one article published by 2 authors, 4 authors, and 8 authors.

Concerning projects and programs, only 03 authors (Knöchelman 2019; Steinerova 2018; Baum 2017) do not use projects or programs in Digital Humanities for their discussions. However, they present relevant questions regarding the themes of digital humanities and open science.

3.1 Text Mining

The texts were gathered by language. In this way, 4 sets of texts were obtained in .txt files: German (brought together in a single document the 5 texts in German); Spanish (only a text in Spanish); English (brought together the 5 texts in English) and Portuguese (brought together the 2 texts in Portuguese).

Table 3. Authorship and projects related to digital humanities

Article	Year	Authors	Project/Program
01	2020	Rollo, MF	Program Memória de Todos
02	2020	Ferla, LAC; Lima, LFS; Feitler, B	Projeto Implementação da tecnologia de Sistemas de Informações geográficas (SIG) em investigações históricas (2012-2013); Projeto Pauliceia 2.0; Projeto de avaliação e edição de verbetes da Enciclopédia Virtual Wikipédia
03	2019	Lahti, L; Marjanen, J; Roivainen, H; Tolonen, M	Bibliographic data science
04	2019	Knöchelmann, M	
05	2019	Anglada, LM	Collection of Museu do Prado; Transcribe Betham; eBird e Mapa Literali Catalã
06	2019	Stigler, J; Klug, HW	Project Konde (acronym for Kompetenznetzwerk Digitale Edition)
07	2018	Hagmann, D	PHAIDRA repository and excavation "Molino San Vincenzo"
08	2015	Blumesberger, S	Projekts e-Infrastructures Austria
09	2014	Wells, JJ; Kansa, EC; Kansa, SW; Yerka, SJ; Anderson, DG; Bissett, TG; Myers, KN; Demuth, RC	The Digital Index of North American Archaeology (DINAA)
10	2019	Wuttke, U; Spiecker, C; Neuroth, H	Project PARTHENOS (acrônimo de Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies)
11	2018	Steinerova, J	
12	2017	Bassett, S; Di Giorgio, S; Ronzino, P	Project PARTHENOS
13	2017	Baum, C	

To identify relevant information in the texts, it was decided to analyze them using the text miner Sobek. However, the system is adapted only to the English and Portuguese languages. When applying the German and Spanish texts, the results were inconsistent since the most frequent words considered by the system were connective.

For this reason, we sought another text mining tool that was open access and that allowed text mining in other languages. Thus, the tool chosen was Voyant Tools.

When applying the English texts in the Sobek software, the following result was obtained:

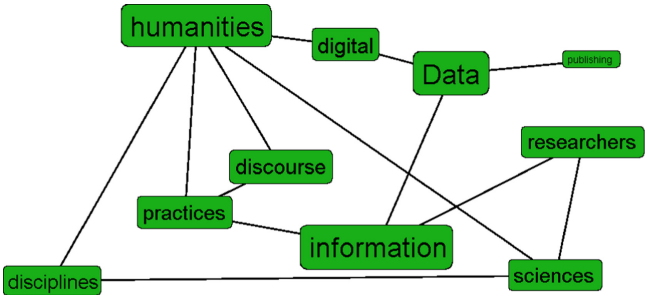


Fig. 3. Sobek mining result English – 15 results

It is possible to observe that the most prominent words are Data, Humanities, and Information. The combination of terms emphasizes the discourse and practices between these elements and the other agents involved, such as disciplines, sciences, researchers, digital (which can refer to the object, the environment, among others), and publication.

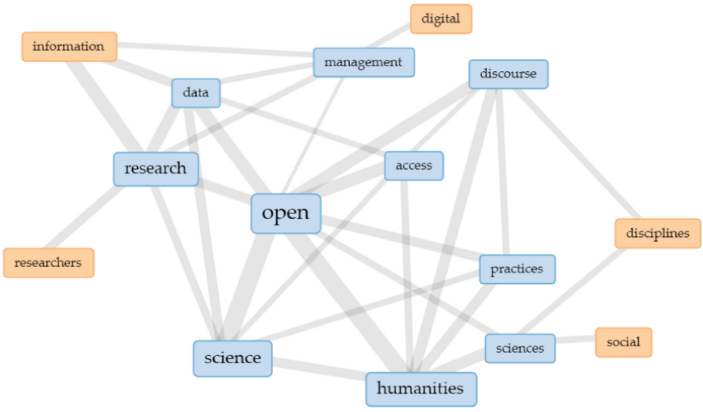


Fig. 4. *Voyant Tools* English mining result – 15 results

In Voyant Tools, it is observed that the keywords are represented in blue. The most relevant word is open (with 255 occurrences), and it has a strong connection (represented by the thickest line) with the terms science (182) and humanities (170). Between the term research (156) and information (101), there is a stronger link, and with the term researchers (43), the link is a little weaker. It is observed that the terms information, researchers, digital, disciplines, social are presented in orange; that is, they are close words, which the system calls co-occurrence.

An important piece of information evident in Voyant Tools and that did not appear in Sobek was the link between the terms Open and Science, which is one of the objects of this research.

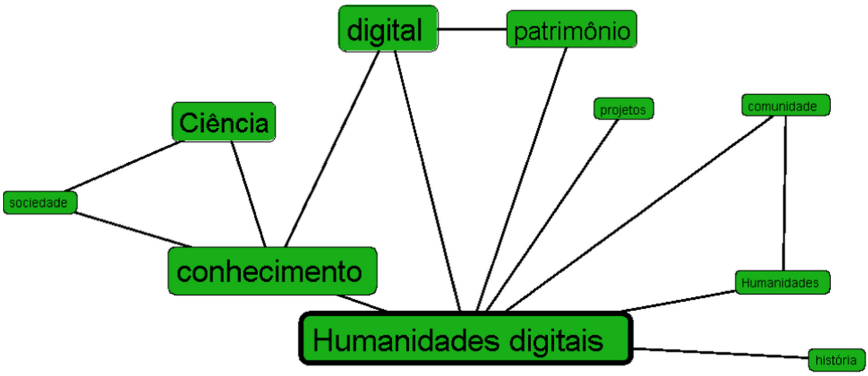


Fig. 5. Sobek Portuguese mining result – 15 results

In Fig. 5, which represents the mining of the Portuguese text in the Sobek software, the prevalence of the term digital humanities and its connection with the terms heritage and digital and the terms knowledge and digital, are observed. The term knowledge also joins science and society. Another link that can be observed is between the digital humanities, the humanities, and the community. The terms project and history have isolated links to the term digital humanities.

The mining result on Voyant Tools (Fig. 6) showed two groups of results. The keyword knowledge is linked to similar terms, such as production, sharing, creation, and scientific. The link between human and digital keywords is strong, and this link is surrounded by the words close together: are, axes, field, scientific, and community. Thus, the system's efficiency is questioned since co-occurrences are and axes are not relevant to the results.

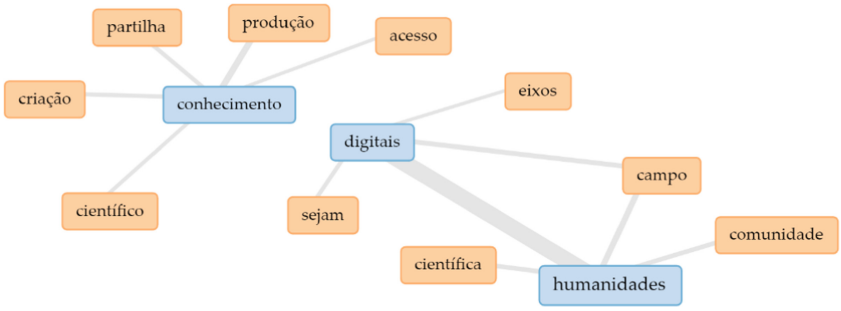


Fig. 6. Voyant Tools mining result Portuguese – 13 results

Figure 7 results in mining in German-language texts. Again, there is a certain fragility since the keywords digitalen and digitale represent the term "digital" and have a connection with digital co-occurrence, which has the same definition. However, most of the other co-occurrences related to digital are tecnologien (technologies), zeitalter

(time), geisteswissenchaften (humanities), verfahren (process/procedure), forschungsinfrastrukturen (research infrastructures) and literaturwissenschaft (literary studies) are relevant to the studies.

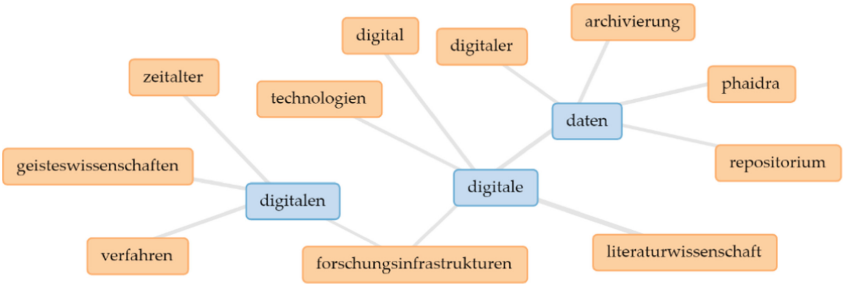


Fig. 7. Resultado mineração *Voyant Tools* Alemão – 14 resultados

Another keyword is daten, which means data. This is linked to digitaler (digital), archivierung (archiving), Phaidra (an acronym for Permanent Hosting, Archiving, and Indexing of Digital Resources and Assets), repositorium (repository) and are aligned with the studies in question.

The German text was also applied to the Sobek miner, but the result was even more inconsistent than found on Voyant Tools since the miner is only enabled for English and Portuguese.

Figure 8 shows the mining result in Voyant Tools for the text in Spanish. However, just as in Sobek, the result using this tool is inconsistent because the three keywords two - "no" and "a" - do not bring a relevant meaning, as well as the co-occurrences linked to them, which were sólo (only, unique), más (more) and hay (has).

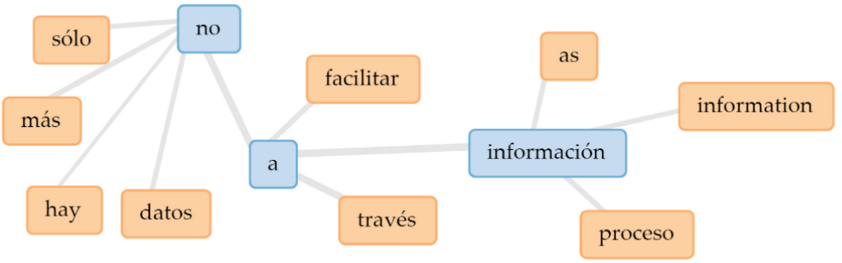


Fig. 8. Spanich Voyant Tools mining result – 12 results

For this reason, it was decided to use the word cloud, which despite including the articles and terms that would normally be excluded, presents the most relevant terms in the text, such as: información (49 ocorrências), ciencia (24), bibliotecas (18), cambios (18) ou cambio (16), digitales (12), abierta (11), investigación (11), centros (10), humanidades (9), among others with less relevance.

Fig. 9. Spanish Voyant Tools word cloud

The results using the Sobek software for the English and Portuguese languages were consistent; however, the same consistency was not maintained for the other languages of the research corpus. The Voyant Tools software, chosen for mining in other languages, also showed inconsistencies, mainly in Spanish.

3.2 Digital Humanities and Actions for Open Science

Among the results found, it was observed that the first researches that address digital humanities and open science are the studies by Wells et al. (2014), in which the project The Digital North American Archeology (DINAA) is presented, which aims to develop models and approaches to publishing and index archaeological data inventoried in historical and archaeological sites in large areas of North America [22].

In 2015, Blumesberger addressed the project to expand repository infrastructure in Austria (Projekts e-infrastructures Austria). The author emphasizes the team's work in selecting objects for scanning, presenting selection reports, defining the necessary metadata, uploading to the repository, and managing the objects until viewing. With the information collected, the team created the Data Management Plan, a document that brings considerations about all phases of the project, from creation to transfer of the data package, also addressing data management in the repository, as well as the reuse and visualization of objects. The group also discusses the use of Linked Open Data (LOD) [23].

Basset, Di Giorgio, and Ronzino (2017) present the PARTHENOS project as a Horizon 2020 project that aims to reinforce research in linguistic studies, cultural heritage, history, archeology, and related areas. To make this possible, PARTHENOS is building an interdisciplinary environment that allows humanities researchers to have access to data, tools, and services based on common policies, guidelines, and standards [24].

In 2019, Wuttke, Spiecker, and Neuroth also brought actions on the Projeto PARTHENOS. According to the authors, the most important areas for the project are to define how long-term archiving will take place, the management of intellectual property, its implementation, the development of common standards, services and methods for interdisciplinary and subsequent use, data use, as well as additional training and education. Throughout the article, the actions developed in work packages are presented, and the most relevant packages are packages 4, 7, and 8. Package 4 has developed a standardization kit consisting of documentation on the relevant human and cultural sciences standards in digital media. This package can also be used for teaching, and in

partnership with packages 7 and 8, develops communication and training materials on the importance of standardization for the research process [25].

In 2018, Hagmann presented the PHAIDRA repository from the University of Vienna, which stores long-term digital data from scientific research in an environment with sustainable standards. Thus, the author considers PHAIDRA ideal for storing archaeological data, such as the data collected at Sítio Molino San Vincenzo in Tuscany, Italy. The University of Vienna has been carrying out archaeological investigations on the site since 2012. There is data collected in digital and analog format. These are scanned for further processing. The author emphasizes the use of information and communication technologies (ICT) in the archelogy, which is a prerequisite for generating digital data in archeology [26].

Returning to the year 2019, Lathi et al. present the project Bibliographic data science. They analyze the science of bibliographic data as a study that derives from the area of data science. The Finnish and Swedish National Bibliographies (FBN and SNB), the English Short-Title Catalogue (ESTC), and the Heritage of the Printed Book database (HPBD) are used for the analysis. These bibliographies cover more than 6 million entries for printed products in Europe and elsewhere. According to the authors of this study, it is possible to demonstrate, through a qualitative approach, the history of the book [27].

According to the authors, efforts for large-scale automated harmonization can improve overall reliability and commensurability between metadata collections, complementing LOD and other technologies that focus on data management and distribution. Bibliographic data science aims to fill an important gap in the area since many bibliographic metadata present inaccurate entries, collection bias, and missing information.

For Anglada (2019), Digital Humanities are a movement that affects information centers' life and evolution. According to the author, society is approaching when what is not on the network is no longer important, so libraries and documentation centers must shift their attention to the document so that it has easy access. Besides, the author suggests promoting citizen participation, as is the case with the following initiatives: Prado Museum Collection, the Transcribe Betham project, the eBird project and the Literali Catalã Map [28].

The author suggests that it is possible to clearly observe the effects of open science on the Digital Humanities since researchers need digital objects to be accompanied by information about the processes that made them readable by the computer to be shared and reused by different groups later.

Stiegler and Klug (2019) exhibit the Konde Project (an acronym for *Kompetenznetzwerk Digitale Edition*), which defines the prerequisites for establishing a platform digital editions, which aims to develop and preserve cultural heritage. The authors claim that digital editions are a product of the digital discipline of Humanities. Computer-aided methods are used to create, research, and disseminate publications from reliable scientific sources [29].

With the Konde project, Austria intends to become a leader in innovation in digital publishing. To this end, it relies on the participation of researchers and professionals from libraries, archives and museums to strengthen research centers and effect collaboration between participants.

Memória de Todos is the program presented by Rollo (2020) and aims to promote heritage education, literacy, digital skills development, and the democratization of access to historical research tools and promote the collection, sharing, and preserving of memories and testimonies [30].

For the author, the Digital Humanities comprise the use of tools, information research, organization, content storage, or even the programming and use of databases or computing tools in their entirety. However, these skills are scarce in the humanities community itself. Also, according to the author, digital has brought a series of opportunities to expand the contents, especially the immaterial contents that can be registered and preserved for future generations.

The authors Ferla, Lima and Feitler present the Projeto Implementação da tecnologia de Sistemas de Informações Geográficas (SIG) technology in historical investigations (2012–2013) and the Pauliceia 2.0 Project that explore the possibilities of geotechnologies in historical investigations. These studies were developed by Grupo de Pesquisas Hímaco – História, Mapas e Computadores and had the participation and support of the Núcleo de Acervo Cartográfico do Arquivo Público do Estado de São Paulo (Apesp) [31].

In addition to the previous projects, there was also the Project's development for the evaluation and editing of entries in the Wikipedia Virtual Encyclopedia, carried out during the disciplines of Modern History I and II of UNIFESP, in which it was proposed to change the entries in Wikipedia. This project supported the Brazilian team linked to the Wikipedia Foundation, which offered lectures and provided guidance on the first steps for editing the content.

To include the Pauliceia 2.0 project within the scope of open science, the project was presented in 2017 in the auditorium of the Archive of the State of São Paulo and proposed to discuss, receive criticisms and suggestions. Besides, this moment was used to request empirical material to support the computational codes' tests. In October 2018, a new presentation was made to present the platform's beta version, in which the community was invited to help with the tests.

The authors Baum (2017), Steinerova (2018), and Knöchelmann (2019) that address Digital Humanities and Open Science but do not use projects for the foundation will be addressed from this point on.

Baum (2017), in your article *Digital gap or Digital turn*? It addresses literary studies and the digital age. For the author, digital has brought a series of advantages since, in digital environments, codes, scripts, and annotation decisions can be disseminated and even discussed and revised more easily through versioning. However, the author also brings up crucial issues such as the gap between first and third-world countries caused by technologies and socioeconomic issues, such as ethnicity, gender, nationality, and education, which include access to digital literacy [32].

In the context of open science, Baum (2017) states that the ability to connect research itself does not prove that everything presupposes digital processes or that everything is subordinate. Mainly because there are some issues imposed by digital that are difficult to reflect in the analog environment due to its high technical complexity, which requires digital expertise and collaborative work in this sector. According to the author, the term Open Science groups strategies and processes that aim at the digitalization opportunity

to make the components of the scientific process accessible and reusable, that is, to bring new opportunities for science, society, and business.

Baum (2017) notes that literary studies are a major challenge for open science. It is a vast and heterogeneous field with various authors and several individual researchers, working groups, institutions, funding agencies, publishers, internet providers, and the networked public. Thus, open access to specialized scientific literature should be allowed, ensuring that it can be referenced and cited, in addition to promoting new forms of specialized communications and the recognition of collaborative research and publications.

In the article *Perceptions of the information environment by researchers: a qualitative study* (2018), Jela Steinerova seeks to understand how to research and academic practices represent a challenge for improving information services and information infrastructures. In her research, the author sought to identify the perceptions and uses of open access and open science resources by researchers. Steinerova identified that researchers consider open access advantageous, mainly due to the increase in citations and publication speed. Still, they also express concerns about commercial influences and evaluations of digital publications [33].

Steinerova also noted that many researchers agree with European open access policies, while others fear the lower quality of digital publishing. Transparency and open access were identified as factors of open science and participation, collaboration, peer networking, and information sharing. Technological determination has been identified in the big data sciences, such as astrophysics, physics, genetics, and others. In the Humanities, the development of digital libraries and archives in cultural heritage was observed. Besides, other open science factors were mentioned, such as policies, evaluation of results, access to data, and publication.

According to Steinerova's research, researchers' social networks share data, information, and publications. Open science was perceived as an advantage, especially open access sources, in addition to interdisciplinary cooperation and advertising. Still, gaps were also observed in the coordination of open science and access to publications and data and concerns about commercial influences and access to finance.

Marcel Knöchelmann, in his article, *Open science in the humanities, or: Open humanities?* (2019) states that open science has deficiencies in addressing the humanities, so it is necessary to think about a discourse on open humanities. For the author, the arguments for the need for this discourse are: the humanities are a by-product of open science, as they do not have their own discourse; the fragmentation of discourses about open practices in humankind requires a unification of these discourses and, mainly, the inadequacies of current scientific communication practices, since there are differences in the communication practices of scientific and human disciplines [34].

According to Knöchelmann the term open humanities has been used previously, but that does not mean an open humanities discourse. For the author, the importance of an open humanities discourse brings together thinking that the humanities need to be open to the transfer of interdisciplinary knowledge, especially concerning digital humanities; Humanities also need a transdisciplinary space in which to shape their digital and open future, working to open up their practices, problems, and implementations. Another

issue pointed out by Knöchelmann concerns the use of copyright licenses since they are important for the authorship's progressive understanding.

3.3 Emphasis on Open Science

Some articles in the corpus emphasize more than others some characteristic aspects of open science. These aspects will be addressed in this session.

Basset, Di Giorgio, and Ronzino (2017) emphasize that because it is a project of Program Horizon 2020, data management should be concerned with meeting the FAIR principles (Findable, Accessible, Interoperable, Reusable); that is, the data must be traceable, accessible, interoperable and reusable. Thus, the projects financed by the Program must implement a Data Management Plan (PGD) to improve and maximize access and reuse of the research data generated by the project. What reinforces the actions of the European Union for the circulation of knowledge.

The authors state that the PGD PARTHENOS model is divided into three levels: the first level includes a set of essential general requirements, regardless of the discipline; the second level includes specific requirements, and the third level is project-based. To assist in completing the PGD, a set of instructions for specific disciplines will be provided. According to the authors, the PGD PARTHENOS model allows researchers to freely access, mine, explore, reproduce and disseminate their data and identify the tools needed to use raw data and validate research data, or to provide their own tools, taking a significant step towards the realization of open science.

Wuttke, Spiecker, and Neuroth (2019) emphasize the FAIR principles and the values of Open Science, increasing the cooperation between the existing research infrastructures, stimulating the exchange of both technical and semantic standards, and expanding the disciplinary boundaries, making that the discussions come out of the walls of the scientific communities and start to involve representatives of data centers, memory institutions, and research associations. According to the authors, these actions are an important contribution to the European Open Science Cloud (EOSC), which aims to be an environment in which researchers can store, analyze and reuse data for research, education, among other purposes. others.

According to Hagmann (2018), PHAIDRA is based on the Fedora Commons Repository and presents information on authorship, licensing, historical research framework, among others. In addition to including more detailed versioning, classification and categorization information. All objects and collections have persistent identifiers (permanent link, handle, and DOI), making it easier to quote. All of these actions allow data to be reused. Besides, data records receive a Creative Commons Attribution 4.0 International (CC BY 4.0) license concerning free and open access whenever possible.

For Lathi et al. (2019), LOD represents a crucial step in making the most of digital resources by integrating web sources and open and reusable metadata.

According to Anglada (2019), the Open Science movement makes science faster, more accurate, and reusable. For the author, there are three criteria for making Open Science, it must be open, collaborative, and made with and for society. According to the author, what is observed is that the Open Science and Digital Humanities movement are a reflection of the profound changes in the investigation and research processes. Thus, it is not a matter of making more scientific dissemination than establishing new

relationships between people and science. Citizens can not only be interested in science, but they can also contribute with their participation and contributions.

Stiegler and Klug (2019) state that the KONDE project has a consistent orientation towards Open Science (Open Access, Open Data, and Open Source). The project also guarantees long-term and free access to research data and allows laypeople to be actively involved, if necessary, and scientifically, if indicated, in the sense of addressing citizen science.

According to Rollo (2019), open science has enormous potential for transformation, especially concerning the Digital Humanities work, since it widens the interconnections between the humanities and society. According to the author, open science focuses not only on open access, open innovation, citizen science, but also on the challenges of archiving and storage, preservation and curation of data, and digital information produced on a large scale society.

3.4 Challenges for Open Science

In 2014, Wells et al. emphasized that the financial issue is a source of concern, as the financing of public goods, such as open data, usually requires public or philanthropic support. The authors also point out that the project's financial sustainability requires lawmakers to understand that databases are America's national heritage.

Another challenge mentioned by Wuttke, Spiecker, and Neuroth (2019) is the large volume of data, tools, and digital methods that have not yet been adequately addressed by scientific communities. The authors also emphasized that communities and project participants have deficits regarding the FAIR principles and the EOSC concept.

Hagmann (2018) describes as a challenge the great diversity of data and obsolete supports. Thus, the challenge is to make these archaeological data available sustainable through effective long-term archiving. For this, the data must be saved in formats suitable for archiving, which allows the development of future research and the preservation of this data.

Lathi et al. (2019) emphasize that the lack of open data availability is a major bottleneck for bibliographic data science's transparent and collaborative development. Furthermore, they criticize, in a way, the term open science because, for the authors, this terminology is not concerned with the humanities but with a grouping of scientific disciplines.

4 Conclusion

Although the Budapest Declaration dates back to 2002 and the open science movement was driven by it [35], publications were found that relate Digital Humanities to open science only from 2014. For the volume of publications only in 2019, greater expressiveness was observed in the number of publications.

The languages with the largest number of publications are English and German, with 5 publications, and Austria is the most productive country with 3 publications. Regarding Qualis Capes, of the 6 journals that appear on the Qualis list, 4 fall into stratum A.

It was hoped that the use of the text miner would show the relationship between digital humanities and open science, but this was not the case. In the analyzed corpus (groupings of texts in German, Spanish, English and Portuguese) many times not even the relationship between the humanities and the digital was evident. This may have occurred because the number of keywords retrieved was small (around 15 results) or due to some other unidentified error in the tool.

Regarding open science findings, the importance of the Data Management Plan (PGD) became evident. For a PGD to be well prepared, there must be documents that guide the researcher, in the different phases of each project, regarding the preparation of a PGD for each area of knowledge, since each area has different specificities.

The standardization of metadata, the use of software that allows versioning of research data, the adoption of permanent links, the attribution of Creative Commons licenses, and, above all, thinking that the data generated must meet the FAIR principles, or that is, data needs to be localizable, accessible, interoperable and reusable, facilitating and making open science possible.

Besides, the creation and development of communication materials, the provision of training aimed at training researchers and citizens, and the collaboration of professionals from different institutions with different backgrounds combined with citizens' participation create a favorable environment for the valorization of open science.

The valorization of open science extends to the spaces that this physical and digital heritage occupies. It expands the importance and gives visibility to the spaces for guarding and preserving memory, such as libraries, archives, and museums. Spaces for education, property protection, democratization, collection, sharing, and memory preservation. Essential spaces for the humanities and the advancement of science.

References

1. de Sousa, M.C.P.: As humanidades digitais globais: Anotações. https://humanidadesdigitais.org/hd2015/anotacoes/. Accessed 13 abr 2019
2. Guerreiro, D., Padre Roberto Busa, S.J.: Bibliotecas e Humanidades Digitais. https://bdh.hypotheses.org/tag/index-thomisticus. Accessed 01 June 2020
3. Dalbello, M.: A genealogy of digital humanities. J. Doc. **67**(3), 480–506 (2011)
4. de Almeida, M.A., Damian, I.P.M.: Humanidades digitais: um campo praxiológico para mediações e políticas culturais. In: XVI Encontro Nacional de Pesquisa em Ciência da Informação. João Pessoa, PB (2015)
5. Sula, C.A.: Digital humanities and libraries: a conceptual model. J. Library Admin. **53**, 10–26 (2013)
6. Burghardt, M., Wolff, C., Womser-Hacker, C.: Informationswissenschaft und digital humanities. Inf. Wissenschaft Praxis **66**(5–6), 287–294 (2015)
7. Koltay, T.: Library and information science and the digital humanities: perceived and real strengths and weaknesses. J. Doc. **72**(4), 781–792 (2016)
8. Pimenta, R.M.: Por que humanidades digitais na ciência da informação? Perspectivas pregressas e futuras de uma prática transdisciplinar comum. Informação e Sociedade: Estudos **30**(2), 1–20 (2020)
9. Puerta Diaz, M., Bisset Alvarez, E., Marín-Arraiza, P., Dutra, M.L.: Análisis de redes sociales en laboratorios de Humanidades Digitales en Brasil: desarrollo y desafíos de la producción científica. In: X Encuentro Internacional de Investigadores y Estudiosos de la Información y la Comunicación, Cuba (2019)

10. Puerta Diaz, M., Bisset Alvarez, E., Vidotti, S.A.B.G.: Humanidades digitais: visualização da produção científica. In: II Workshop de Informação, Dados e Tecnologia (WIDaT 2018), João Pessoa (2018)
11. Tang, M., Cheng, Y.J., Chen, K.H.: A longitudinal study of intellectual cohesion in digital humanities using bibliometric analyses. Scientometrics **113**(2), 985–1008 (2017)
12. dos Santos, P.X., de Almeida, B.A., Henning, P.: Livro verde - ciência aberta e dados abertos: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional. Fiocruz, Rio de Janeiro (2017)
13. Albagi, S., Maciel, M.L., ABDO, A.H.: Ciência aberta em questão. In: Ciência aberta, questões abertas. Brasília, IBICT (2015)
14. Packer, A.L., Cop, N., Santos, S.M.: A rede SciELO em Perspectiva. In: Packer, A.L., Cop, N., Luccisano, A, Ramalho, A., Spinak, E. (eds.) SciELO – 15 anos de acesso aberto: um estudo analítico sobre acesso aberto e comunicação científica. UNESCO, Paris (2014)
15. Atlas, T.I.: Huge Variety Of Media Types. https://atlasti.com/product/what-is-atlas-ti/. Accessed 29 Jan 2021
16. GTECH.EDU. Mineração de texto educacional, http://sobek.ufrgs.br/uploads/sobek_quick_reference_guide_pt.pdf. Accessed 30 Sept 2020
17. LARHUD. Voyant Tools. http://www.larhud.ibict.br/index.php?title=Voyant_Tools. Accessed 27 Sept 2020
18. ERCIM.EU. ERCIM Flyer. https://www.ercim.eu/download/ERCIM-flyer-web.pdf. Accessed 04 Sept 2020
19. WEB OF SCIENCE GROUP. Master journal list. https://mjl.clarivate.com/home. Accessed 02 Oct 2020
20. SCOPUS. Sources. https://www-scopus-com.ez22.periodicos.capes.gov.br/sources.uri. Accessed 19 Aug 2020
21. CAPES. Critérios de classificação Qualis – Ensino. https://www.capes.gov.br/images/stories/download/avaliacaotrienal/Docs_de_area/qualis/ensino.pdf. Accessed 21 Aug 2020
22. Wells, J.J., et al.: Web-based discovery and integration of archaeological historic properties inventory data: the Digital Index of North American Archaeology (DINAA). Literary Linguist. Comput. **29**(3), 349–360 (2014)
23. Blumesberger, S.: Der welt der metaddaten im universum von repositorien. Mitteilungen der VÖB **68**(3/4), 515–528 (2015)
24. Bassett, S., Di Giorgio, S., Ronzino, P.A.: Data management plan for digital humanities: the PARTHENOS Model. ERCIM, Special Theme (111) (2017)
25. Wuttke, U., Spiecker, C, Neuroth, H.: PARTHENOS – eine digitale Forschunsinfrastruktur für die Geistes- und Kulturwissenschaften. Bibliothek – Forschung und Praxis **43**(1), 11–20 (2019)
26. Hagmann, D.: Überberlegungen zur nutzung von PHAIDRA als repositorium fü r digitale archäologische daten. Mitteilungen der VÖB **71**(1), 53–69 (2018)
27. Lathi, L., et al.: Bibliographic data science and the history of tha book (c, 1500–1800). Cataloging Classif. Q. **57**(1), 5–23 (2019)
28. Anglada, L.M.: Muchos cambios y algunas certezas para las bibliotecas de investigación, especializadas y centros. El profesional de la informacion **28**(1) (2019)
29. Stiegler, J., Klug, H.W.: KONDE – Ein netzwerk bringt forschuns – und GLAM-institutionen zusammen. Ein projektbericht. Mitteilungen der VÖB **72**(2), 431–439 (2019)
30. Rollo, M.F.: Desafios e responsabilidades das humanidades digitais: preservar a memória, valorizar o patrimônio, promover e disseminar o conhecimento. Programa Memória para todos. Estudos Históricos **33**(69), 19–44 (2020)
31. Ferla, L.A.C., Lima, L.F.S., Feitler, B.: Novidades no front: experiências com humanidades digitais em um curso de história na periferia da grande São Paulo. Estudos Históricos **33**(69), 111–132 (2020)

32. Baum, C.: Digital gap oder digital turn? Literaturwissenschaft und das digitale Zeitalter. Zeitschrift für Germanistik **27**, 316–328 (2017)
33. Steinerova, J.: Perceptions of the information environment by researchers: a qualitative study. Inf. Res. **23**(4), (2018)
34. Knöchelmann, M.: Open science in the humanities, or: open humanities? Publications **7**(65), 1–17 (2019)
35. Dos Santos, P.X., de Almeida, B.A.: HENNING, p. 11 (2017)