

# ESTUDO CIENTOMÉTRICO DA ATIVIDADE ACADÊMICA SOBRE AS TEMÁTICAS DE HUMANIDADES DIGITAIS E BIG DATA NAS UNIVERSIDADES ESTADUAIS PAULISTAS<sup>1</sup>

Scientometric study of scholarly activity on the themes of digital humanities and big data at São Paulo state universities

**Guimarães, Suzane Albino da Silva**

Universidade de São Paulo,

Departamento de Informação e Cultura, Escola de Comunicações e Artes, São Paulo, SP, Brasil  
suzane.albinoguimaraes@gmail.com

<https://orcid.org/0000-0002-3432-6434> 

**Rocha, Ednéia Silva Santos**

Universidade de São Paulo,

Departamento de Educação, Comunicação e Informação,  
Ribeirão Preto, SP, Brasil  
edneia@usp.br

<https://orcid.org/0000-0003-1478-6828> 

**Mugnaini, Rogério**

Universidade de São Paulo,

Programa de Pós-Graduação em Ciência da Informação,  
Escola de Comunicações e Artes, São Paulo, SP, Brasil  
mugnaini@usp.br

<https://orcid.org/0000-0001-9334-3448> 

A lista completa com informações dos autores está no final do artigo 

## RESUMO

**Objetivo:** Mapear a atividade acadêmica sobre as temáticas de Big Data e Humanidades Digitais de pesquisadores da Universidade de São Paulo (USP); Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP) e Universidade Estadual de Campinas (UNICAMP).

**Método:** abordagem quantitativa, sendo de natureza aplicada através de um estudo cientométrico. Quanto aos objetivos, trata-se de uma pesquisa exploratória, e quanto aos procedimentos, bibliográfica e documental. O núcleo da pesquisa constitui-se das atividades acadêmicas sobre Big Data e Humanidades Digitais, de pesquisadores das universidades estaduais paulistas recuperadas: na base de dados Web of Science (publicações); na Biblioteca Virtual da FAPESP e do DGP/CNPQ (auxílios à pesquisa), e; no Currículo Lattes dos pesquisadores identificados nas bases da FAPESP e CNPq (atividades acadêmicas).

**Resultados:** apontam que as temáticas se concretizam em publicações na Web of Science a partir de 2013 com um total de 281 publicações, sendo 92,9% de Big Data. A USP está presente em 64,1%, a UNICAMP em 25,6% e a UNESP em 10,03%. As principais áreas do conhecimento relacionadas com as temáticas foram, respectivamente, as Ciência da Computação, Medicina e Ciência da Informação.

**Conclusões:** possibilitou a compreensão de que a escolha por uma metodologia de pesquisa que inclui, tanto as pesquisas realizadas (publicações) como as outras atividades acadêmicas como os auxílios de pesquisa (projetos (pesquisa potencial) e bolsas), pode oferecer um olhar abrangente daquilo que está se desenvolvendo e se concretizando sobre essas temáticas nas universidades estaduais paulistas. Revelando, portanto, que um estudo limitado a Web of Science representaria principalmente as temáticas nas áreas das Ciências Exatas e da Terra. A maior abrangência das fontes adotadas (visualizada nos auxílios de pesquisa FAPESP e CNPq) permitiu observar o papel que as Humanidades e Ciências Sociais Aplicadas vêm desempenhando, e mais importante, a inserção da Ciência da Informação.

**PALAVRAS-CHAVE:** Humanidades Digitais; Big Data; Estudo cientométrico; Atividade acadêmica; Universidades estaduais paulistas.

## ABSTRACT

**Objective:** To map the scholarly activity on Big Data and Digital Humanities research topics at the University of São Paulo (USP); São Paulo State University “Júlio de Mesquita Filho” (UNESP) and State University of Campinas (UNICAMP).

<sup>1</sup> Uma versão preliminar deste estudo foi apresentada no 8º Encontro Brasileiro de Bibliometria e Cientometria, celebrado em julho de 2022 na UFAL, Maceió-AL.

**Methods:** analytical approach, being applied through a scientometric study. As for the objectives, it is an exploratory research, and as for the procedures, bibliographic and documentary. The core of the research is recovered from the scientific activities on Data and Humanities constituting São Paulo: in the Big Web of Science (publications); in the FAPESP and DGP/CNPq Virtual Library (research grants), and; in the Lattes Curriculum of the researchers identified in the FAPESP and CNPq databases (scientific activities).

**Results:** They point out that the themes materialize in publications on the Web of Science from 2013 onwards, with a total of 281 publications, 92.9% of which are Big Data. A USP is present in 64.1%, a UNICAMP in 25.6% and a UNESP in 10.03%. As main areas of knowledge related to thematic areas such as Computer Science, Medicine and Information Science.

**Conclusion:** it made it possible to understand that the choice of a research methodology that includes, both as carried out (publications) and other scientific activities such as research grants (projects (potential research) and grants), can offer a comprehensive look. establishing and materializing on universities in state universities. Revealing, therefore, that a study limits the representation of the Web of Science mainly as a theme in the areas of Exact and EarthSciences. The wider range of sources adopted (visualized in the FAPESP and CNPq research grants) allowed us to observe the role that the Humanities and Applied Social Sciences have been playing, and more importantly, the insertion of Information Science.

**KEYWORDS:** Digital Humanities; big data; Scientometric study; Scientific activity; Paulista State University.

## 1 INTRODUÇÃO

Na sociedade da informação o grande volume de dados intitulado Big Data (BD) tem multiplicado as formas de organizar e analisar as informações, relacionando diferentes profissionais, como os cientistas da computação ou cientistas de dados, a diversas áreas, com as quais tradicionalmente não possuíam uma relação direta.

Esta aproximação e interação entre as comunidades das Ciências Humanas, Sociais e Sociais Aplicadas (CHSSA) com os cientistas da computação deram origem a um campo interdisciplinar nomeado de Humanidades Digitais (HDs). Esse termo é utilizado para definir a pesquisa que incorpora a aplicação de computação e Tecnologia da Informação e Comunicação (TICs) a estudos das CHSSA, como também a estudos das CHSSA que investigam as influências das TICs na sociedade e na cultura.

As HDs se caracterizam como um campo científico essencialmente colaborativo, que se utiliza de diversas ferramentas e métodos de pesquisa para analisar grandes conjuntos de dados que não poderiam ser mensurados sem o apoio de ferramentas computacionais. Esses dados incluem a digitalização de acervos de bibliotecas e museus e a criação de conjuntos gigantes de registros digitais. Assim, numa nova cultura digital, grandes quantidades de dados serão geradas diariamente, por meio de blogs, comentários, e-mails, pesquisas na Internet, downloads de músicas, "compartilhamentos" e compras online. Exigindo, portanto, novas ferramentas analíticas para identificar tendências e padrões, e extrair informações importantes. Essa expansão da tecnologia fornecerá insights sobre as tendências culturais, artísticas e intelectuais do passado que podem ser digitalizados, ampliando o escopo de disseminação da informação no contexto digital (CHEN, 2015).



Nesse cenário, Kaplan (2015) se detém à pesquisa de BD em HDs, caracterizando um ponto de interseção – que seria parte do domínio maior, ou seja, as HDs –, ilustrando-o por meio de três círculos concêntricos, onde: o primeiro

, que seria o núcleo, corresponde aos grandes conjuntos de dados culturais, cujo processamento e interpretação são o desafio patente; o segundo estaria relacionado à necessidade de se compreender a relação entre discurso coletivo, comunidades gigantescas, e softwares mediadores e respectivos algoritmos, no contexto do que se chamaria Cultura Digital; e o terceiro circunda os anteriores, se atendo à experiência humana na lida com tais conjuntos de dados, através de vários tipos de interfaces imersivas (por exemplo, permitindo a transição de uma experiência mais abstrata no Google Maps, a um ambiente mais completo e realista, como o Google Street View).

A presença de pesquisadores de CHSSA na produção científica em BD pode estar relacionada à antecedência do campo de “computação das humanidades” (*computing humanities*) cuja origem remonta à metade do século XX, e na virada do século é considerado sob o termo guarda-chuva das HDs. Pode-se então afirmar que HDs e BD e são contemporâneos (SVENSSON, 2009), resultando nos primeiros artigos na Web of Science (WoS), respectivamente, de 2004<sup>2</sup> e 1998 (HU; ZHANG, 2017; SU; ZHANG; IMMEL, 2020). Por outro lado, a produção científica brasileira em tais temáticas não evidencia uma tendência similar<sup>3</sup>, além de se dar mais recentemente, exigindo um olhar mais amplo do ciclo de vida das pesquisas. Daí entende-se a necessidade de mirar esforços que podem anteceder as publicações (pesquisa realizada) (DAVYT; VELHO, 2000), e que ofereçam informações da atividade acadêmica – como projetos de pesquisa, organização de eventos, participação em bancas, orientações em pesquisa, dentre outras.

A abordagem adotada neste estudo visa o monitoramento dessas temáticas recentes marcados pela interdisciplinaridade, com um olhar que não se limita às lentes da produção científica (por meio das publicações). Nesse sentido, propõe-se uma análise

---

<sup>2</sup> Diferente de Hu e Zhang (2017), Lyu e Costas (2021) identificam o ano de 1993 como o mais recente. Contudo uma busca na WoS evidencia que os artigos de anos anteriores a 2004 apenas mencionavam termos como “big dataset”, sem qualquer vinculação com o termo “big data”, que se consolidaria posteriormente. Hu e Zhang (2017) revisaram os resultados e decidiram restringir a busca do termo somente ao título ou palavras-chave do artigo (ignorando o resumo), garantindo que a pesquisa tratasse efetivamente de Big Data.

<sup>3</sup> Os registros com presença de ao menos um autor com afiliação brasileira na WoS apresentam tendência contrária, sendo o mais antigo de BD (2012, com mais de 1200 registros) e o de HDs é de 2014 (com apenas 61 registros).

cientométrica, por analisar quantitativamente a atividade acadêmica em temáticas específicas.

Dando um passo atrás, analisa as atividades acadêmicas empreendidas por pesquisadores de um dos estados melhor equipados economicamente e principalmente no que diz respeito à infraestrutura de pesquisa. Questiona o quanto as referidas temáticas têm sido objeto de interesse das atividades nas universidades estaduais paulistas (UEPs), a saber: Universidade de São Paulo (USP); Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP) e Universidade Estadual de Campinas (UNICAMP).

Considerando as UEPs, que contam com a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) – a mais antiga e uma das principais do país –, buscou-se mapear detalhadamente como se dá o esforço para direcionamento e desenvolvimento de pesquisa nas temáticas Big Data e Humanidades Digitais. Na hipótese de que não se observassem projetos financiados pela FAPESP, optou-se por identificar os grupos de pesquisa registrados no DGP/CNPq. Já a produção científica, quando não identificada na Web of Science, recorreu-se ao currículo Lattes, possibilitando assim a presença em veículos não indexados.

## 2 BIG DATA: DEFINIÇÕES E CARACTERÍSTICAS

Big Data é comumente associado a uma grande quantidade de dados sendo considerado também uma técnica, um fenômeno - uma avalanche de dados - no qual um montante de diversos tipos de dados, por vezes desordenados, registros do cotidiano como textos, áudios, vídeos, redes sociais, dados sobre o consumo, entre outros, podem ser analisados, convertidos e utilizados para responder a diversas demandas da sociedade contemporânea em seus aspectos econômicos, sociais e culturais.

A sociedade atual, comumente chamada de sociedade da informação, é proveniente do período pós-guerra e do Vale do Silício, quando houve um aumento da demanda de informação juntamente com o desenvolvimento de TICs que pudessem processar automaticamente essas informações, auxiliando igualmente nos processos comunicacionais (BARRETO, 2014). Essa demanda é proveniente da inovação tecnológica a fim de acompanhar as necessidades da sociedade em relação a diversos nichos, como saúde, economia, indústria bélica, corrida espacial, agronomia, entre outras. Tal demanda também se relaciona com o aumento do número de pesquisas, produções científicas e a



sua disponibilização tornando-se necessário criar mecanismos eficientes para o tratamento e disseminação de tais informações.

Sistemas ágeis para processamento de grandes volumes de dados, que localizam, extraem, organizam e transformam dados em informações úteis, certamente possibilitaram a existência do BD. A revolução das TICs ganhou um lugar de destaque na cronologia dos dados até a atualidade, inicialmente mais focada no aspecto tecnológico, mas atualmente se voltando mais para a informação (MAYER-SCHONBERGER, 2013).

Na década de 1980 o volume de dados crescia exponencialmente, surgindo então dúvidas do que fazer com esses dados armazenados. Na época as técnicas utilizadas já não supriam as necessidades de tratamento dos dados, com essa demanda crescente surge no fim da década de 1980 o conceito e a utilização de mineração de dados, até chegar à ideia de BD, o que ocorreria em 1990 (CALDAS; SILVA, 2016).

Para Lima Junior (2011), Baracho (2015), Kaladhar; Doraswamy Naick e Somasekhara Rao (2018) e Mayer-Schonberger (2013), BD se refere a um *dataset* (conjunto de dados) cujo tamanho está além da habilidade de ferramentas típicas de bancos de dados capturar, gerenciar e analisar, ou seja, esse grande volume de dados produzido não pode ser processado sem estruturas computacionais específicas.

Tanto para Barreto (2014) como para Dutra e Matias (2014) BD é uma volumosa quantidade de informações estruturadas, semiestruturadas, não estruturadas e complexas que cresce e é disponibilizada exponencialmente. Ainda para Dutra e Matias (2014, p. 4250, grifo nosso), há uma dualidade na diferenciação relativa ao conceito de Big Data, ora eles apresentam o termo como volumosas quantidades de *informações*, ora como,

[...] uma vasta quantidade de *dados* disponíveis em diferentes níveis de complexidade – criados por humanos ou por máquinas, em diferentes ritmos – e que apresentam grandes níveis de ambiguidade, de forma que não podem ser processados computacionalmente por meio da utilização de tecnologias, dispositivos de comunicação, métodos de processamento e algoritmos tradicionais, além de qualquer outra solução similar.

Ou seja, para alguns autores o BD é reconhecido como uma volumosa quantidade de *dados* e para outros, volumosa quantidade de *informação*, e mesmo esses termos sendo por vezes usados como sinônimos existe uma distinção entre eles. De forma prática, o campo das TICs define dado como “[...] a menor partícula da informação que, quando isolado, não possibilita a decodificação clara e comum ao entendimento humano, pois está fora do contexto que favorece a sua compreensão” (LIMA; ALVARES, 2012, p. 24). Já a

informação é entendida como um conjunto de dados contextualizados no qual se pode verificar o seu significado.

O BD geralmente é apresentado em cinco dimensões: **volume**, **velocidade**, **variedade**, **veracidade** e **valor**, conhecidos com os 5 Vs. Entende-se que o **Volume** se refere à magnitude dos dados, no qual se assume que é impraticável definir um limite específico para um grande volume de dados, já que as definições de grande volume de dados são relativas e variam de acordo com muitos fatores que incluem o tempo e os tipos de dados (CALDAS; SILVA, 2016). Somado a essa questão da grande quantidade de dados é preciso entender que a maior parte desse montante de dados não é estruturada, estima-se que os dados estruturados, ou seja, prontos para análise pertençam apenas a um pequeno subconjunto do BD (KALADHAR; DORASWAMY NAICK; SOMASEKHARA RAO, 2018).

A **Velocidade** se associa com a melhoria nas transmissões e nas emissões de sinal na comunicação e está diretamente conectada com a agilidade em que os dados são gerados, atualizados, analisados e dispostos para uso, muitas vezes gerando informações no âmbito da competitividade e inovação que podem ser decisivas (MOSTAFA; CRUZ; AMORIM, 2015). A **Variedade** dificulta a organização e análise dos dados e está relacionada com a diversidade da origem dos dados, ou seja, existe uma ampla variedade de formatos provenientes de vídeos, metadados, gifs, áudios, dados de sensores e navegadores, reconhecimento facial, sucessão de cliques em sites, textos, imagens, aplicativos de smartphone, câmeras de monitoramento e monitoramento ambiental, entre outros.

Outra faceta do BD está relacionada com a necessidade de lidar com dados imprecisos e incertos, sendo a **Veracidade** associada com a qualidade e ou confiabilidade dos dados. O **Valor** pode ser tanto associado a dados que gerem resultados (dados valiosos), isto é, que geram informações relevantes e para isso precisam ter qualidade, como também está associado ao retorno dos investimentos que foram locados em todos os outros 'Vs' a fim de atribuir sentido ao volume de dados de maneira estratégica (FAGUNDES; MACEDO; FREUND, 2018). A informação há tempos é considerada valiosa, porém os dados eram vistos como subproduto de determinadas operações, atualmente se observa que o valor da grande maioria dos dados passou do uso secundário para o uso potencial, sugerindo que seu valor é muito maior do que o valor imediato.



De maneira geral todos os ‘Vs’ se relacionam entre si, com exceção do **Valor** e a da **Velocidade** que “[...] possuem relações inversas, pois quanto mais tempo se leva para que os dados sejam transformados em informações relevantes, menos valor terá para o negócio, uma vez que resultados obsoletos prejudicam a qualidade e a rapidez na tomada de decisão” (FAGUNDES; MACEDO; FREUND, 2018, p. 200).

O BD traz à tona a complexidade que está envolvida nos processos e mecanismos utilizados para tratar os dados que geram informação e viabilizam a geração de conteúdos significativos que possibilitem colocar as organizações e as pessoas de maneira positiva na tomada de decisões inovadoras, trabalhando e entendendo melhor o mercado e a sociedade.

### 3 VISÃO GERAL DAS HUMANIDADES DIGITAIS

Os números, os dados e mais atualmente o BD são produtos facilmente manipulados e analisados pelos aparatos tecnológicos mais modernos, porém os textos e seus caracteres podem ser decodificados e quantificáveis até por tecnologias computacionais mais antigas.

As Humanidades Digitais são entendidas como um novo campo de investigação intelectual em expansão, de natureza metodológica e interdisciplinar que realiza uma intersecção entre as tecnologias computacionais e as CHSSA (SANTAREM SEGUNDO, 2015; KIRSCHENBAUM, 2012).

Nas áreas das CHSSA se observa que os processos do fazer acadêmico não necessariamente são completamente dependentes ou subordinados aos aparatos tecnológicos, mas são esses recursos tecnológicos que vem transformando a práxis no campo de conhecimento dessas ciências humanísticas e sociais, fazendo das HDs não somente um conjunto de práticas e sim um novo campo acadêmico que, além de utilizar-se das ferramentas tecnológicas em prol de uma mudança das práticas tradicionais para novas práticas, avalia como o uso dessas afeta as humanidades.

As HDs atualmente são consideradas como um campo científico que visualiza as mudanças que as TICs estão produzindo nos campos da investigação humanista, não sendo consideradas como um subcampo específico que surgiu da computação em humanidades (FITZPATRICK, 2012). Além disso, pode-se associar a este contexto a maneira pela qual as humanidades colaboram no entendimento de questões sociais e



culturais levantadas pelo desenvolvimento de tecnologias digitais. Os estudos nessa área são necessariamente colaborativos, pois englobam habilidades, disciplinas e áreas de especialização.

No aspecto interdisciplinaridade Schreibman; Siemens e Unsworth (2004) dizem que,

[...] a computação em humanidades teve que abraçar ‘as duas culturas’, para trazer o rigor e sistemáticas metodologias processuais inequívocas características das ciências para resolver os problemas dentro das humanidades que até então tinham sido mais frequentemente tratados como uma descoberta acidental.

As HDs emergiram a partir de 1949 evidenciada pelo registro do primeiro trabalho que uniu as Humanidades com a Computação. Tal estudo foi arquitetado pelo teólogo Roberto Busa que, procurando Thomas J. Watson (IBM) pretendia obter ajuda na indexação dos trabalhos textuais de Tomás de Aquino. O trabalho idealizado e desenvolvido por Busa era de caráter qualitativo, já que ele não queria realizar apenas uma contagem das palavras. O resultado dessa parceria foi *O Index Thomisticus* disponível na internet desde 2005 (SCHREIBMAN; SIEMENS; UNSWORTH, 2004).

Em 1966 foi publicado o primeiro periódico relacionado com a temática - *Computers and the Humanities* e entre as décadas de 1970 a 1980 houve uma consolidação da área, verificando-se um crescimento no número de publicações científicas, projetos, e associações como a *Association for Literary and Linguistic Computing* em 1973 (SCHREIBMAN; SIEMENS; UNSWORTH, 2004).

Entre 1980 e 1990 constatou-se que os computadores pessoais começaram a ser introduzidos na vida acadêmica, porém ainda com um alto custo, fazendo com que muitos pesquisadores das humanidades ainda dependessem dos centros de computação das universidades (SCHREIBMAN; SIEMENS; UNSWORTH, 2004). Nesse período a Linguística Computacional já trabalhava com bancos de dados lexicais possibilitando a construção de *corpus* linguístico, porém as outras humanidades ainda não se beneficiavam plenamente dos recursos disponíveis em computação voltados para as humanidades. Esse quadro é alterado a partir de 1990 com a internet, facilitando que a computação em humanidades atingisse um público mais amplo e inserindo no cotidiano das pessoas oportunidades de alcance muito maior do que o imaginável (JONES, 2012; SOUSA, 2014).

A partir dos anos 2000 há o crescimento de um senso de identificação da comunidade científica em relação à temática. Com a publicação em 2004 do *Blackwell*



*Companion to Digital Humanities*, o termo passou a ser mais visado. A ampla aceitação das Ciências Humanas Digitais possibilitou a criação da *Alliance of Digital Humanities Organisations* (ADHO) em 2005 e no ano seguinte, com a alteração do nome da conferência anual da *Association for Computers and the Humanities* (ACH) para *International Digital Humanities Conference*, foi possível verificar o destaque que o termo de HDs conquistava (UNIVERSITY OF CAMBRIDGE, 2020).

Em 2010 na THATcamp Paris foi lançado o Manifesto das Humanidades Digitais, importante instrumento para o reconhecimento e divulgação da área. Em 2011 o grupo de pesquisa sobre Humanidades Digitais da USP publicou que existiam 114 centros de pesquisa em 24 países sobre a temática de HDs (SOUSA, 2011). Em abril de 2019 constavam no *Digital humanities - course registry*, somente no Hemisfério Norte 202 cursos e disciplinas sobre as HDs, sendo 190 localizados na Europa e Ásia, e 12 nos Estados Unidos (DARIAH; CLARIN, [2022]).

No Brasil em 2013, na Biblioteca Brasiliana Guita e José Mindlin da Universidade de São Paulo (USP) aconteceu o 1º Seminário Internacional em Humanidades Digitais no Brasil, que contou com a participação de diversos pesquisadores internacionais da área. Em 2018, a Escola de Ciências Sociais (FGV CPDOC) em parceria com o Laboratório de Preservação e Gestão de Acervos Digitais (LABOGAD/UNIRIO), promoveram o I Congresso Internacional em Humanidades Digitais no Rio de Janeiro (HDRio) que reuniu pesquisadores e interessados em explorar e interrogar a produção, a organização e a difusão da informação no meio digital. A segunda edição do evento ocorreu em 2021 no formato online e há programação e chamada para o próximo que ocorrerá em 2023<sup>4</sup>. Em 2022 foi lançado o site da Associação Brasileira de Humanidades Digitais (ABHD) e a Revista Brasileira em Humanidades Digitais, cujo escopo é publicar estudos teóricos e aplicados no campo transdisciplinar das HDs. Tal trajetória evidencia que as HDs estão se consolidando como área de pesquisa.

### **3.1 Humanidades Digitais e Ciência da Informação**

Visto que o campo das Humanidades Digitais é reconhecido como um ambiente transdisciplinar em pleno desenvolvimento se faz necessário pensar em alguns aspectos

---

<sup>4</sup> III Congresso Internacional em Humanidades Digitais (HDRio2023). Disponível em: <https://www.even3.com.br/hdrio2023/>



semelhantes entre as HDs e a área da Biblioteconomia e CI, já que as HDs vêm utilizando-se das inúmeras estratégias tecnológicas para facilitar a guarda e manipulação dos registros permitindo assim um peculiar processo de troca que permeia as narrativas dos arquivos, das bibliotecas tradicionais e digitais, museus e centros de documentação.

A principal semelhança entre as duas disciplinas está no foco geral de que ambas são baseadas nos estudos e nas práticas de registros de informações, ou seja, documentos em um sentido amplo que são suporte da informação registrada, sendo, portanto, nos ambientes como as bibliotecas, arquivos, centros de documentação e museus onde tais suportes se encontram, favorecendo a colaboração entre as áreas. Outra aproximação se refere ao fato de que ambas estão ligadas aos serviços de computação universitária, como também são reconhecidas por terem disciplinas acadêmicas e técnicas de apoio à pesquisa (ROBINSON; PRIEGO; BAWDEN, 2015).

Cunningham (2010) examinou questões relacionadas ao acesso, preservação e disseminação de conteúdos digitais acadêmicos. Discutiu sobre as parcerias de HDs entre professores e bibliotecários, bem como destacou que os projetos de HDs oferecem oportunidades de colaboração entre bibliotecários, pesquisadores e cientistas sociais.

HDs é um termo que engloba uma série de estudos de humanidades que fazem uso da tecnologia da informação. Esses usos variam entre os campos de estudo das humanidades e podem incluir desde imagens e artefatos digitais a texto codificado. As HDs incluem projetos que buscam criar coleções digitais com o objetivo de forjar novos tipos de pesquisa facilitados pelo acesso e ferramentas digitais. Outro tipo de projeto de HDs pode ser visto no desenvolvimento de bibliotecas digitais à medida que as bibliotecas iniciam projetos de digitalização de seus recursos. (CUNNINGHAM, 2010). Para Moura (2019, p. 60) as HDs “são uma nova forma de se abordar as pesquisas em humanidades e ciências sociais, particularmente atravessadas pelo imperativo tecnológico”. A autora ainda observa sinergia entre a CI e HDs, pois as pesquisas nesses campos são direcionadas para investigação das “condições de produção, coleta e tratamento dos dados e com os impactos epistemológicos decorrentes da automação da pesquisa e do uso de algoritmos e Big Data” (MOURA, 2019, p. 65). No que se refere aos aspectos éticos relacionados às interações dos estudos em CI com as HDs destacam-se diligências para aprimoramento de pesquisas que busquem compreender as perspectivas evolutivas dos campos.

Diante desse cenário, as similaridades são observadas no interesse comum pelo ciclo de vida da informação, por afinidades teóricas, e pelas semelhanças na execução da



cadeia de atividades, as quais há uma preocupação com os processos de pesquisa, recuperação de informação, classificação, gestão da informação e até com a bibliometria (ROBINSON; PRIEGO; BAWDEN, 2015).

Por mais que esses apontamentos evidenciem uma similaridade e aproximação há de se pensar que grande parte das atividades e dos resultados está alocada nos diferentes estágios de uma cadeia de processos, pois enquanto a maior parte do trabalho da CI e Biblioteconomia está vinculado com a gestão, organização e recuperação das informações, as HDs estão mais envolvidas com a criação e o uso das informações como produto.

Arrematando os tópicos discutidos até aqui, se demonstra que as HDs são na verdade as humanidades da atualidade que se configuram como um campo híbrido e complexo, atuante não apenas na pesquisa e extensão, mas também no campo de ensino e, principalmente no âmbito do acesso à informação, contribuindo e alterando a lógica da academia e de seus processos cotidianos ‘do fazer acadêmico’ que está cada vez mais digital.

## 4 PROCEDIMENTOS METODOLÓGICOS

A metodologia da pesquisa caracteriza-se por uma abordagem de cunho quantitativo, através de um estudo cientométrico. Quanto aos objetivos, trata-se de uma pesquisa exploratória, e quanto aos procedimentos, bibliográfica e documental.

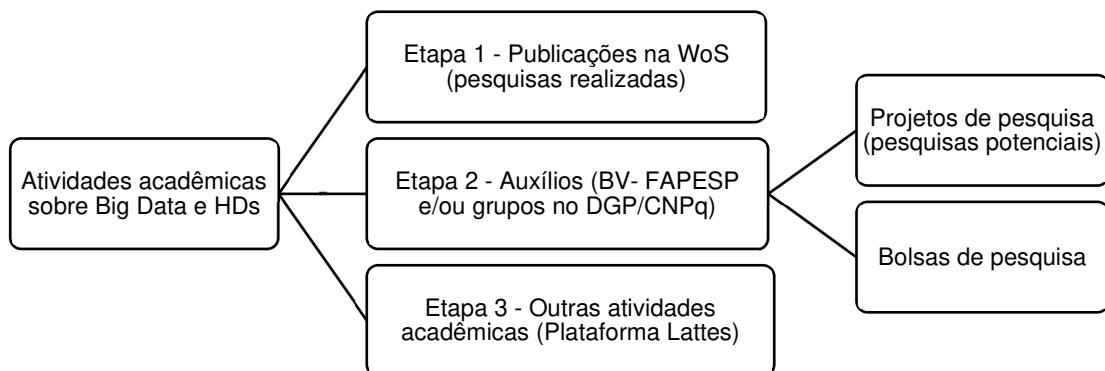
Diante da atualidade das temáticas Big Data e Humanidades Digitais na pesquisa brasileira e ainda sua repercussão internacional, viu-se a necessidade de ampliação do escopo de análise, abarcando as atividades acadêmicas e não se restringindo às publicações. Para tanto, recorreu-se à Cientometria, a fim de se explorar o desenvolvimento da pesquisa nessas temáticas de maneira mais abrangente: desde sua concepção até sua publicação, mas também abarcando outras atividades a ela associada.

Davyt e Velho (2000) definem a ‘pesquisa realizada’ (ou demonstrada) a que se caracteriza pelo ato de tornar público um conhecimento, sendo necessário que o trabalho científico em questão, passe do domínio privado para o público (seja publicado). Já a ‘pesquisa por fazer’ (ou potencial) refere-se aos projetos e bolsas de pesquisa com financiamento de agências de fomento (ou governos e universidades), que através do apoio ao pesquisador propiciam a execução de pesquisas (ou sua concepção) que façam a vinculação entre ciência e desenvolvimento. Somam-se a estas, diversas atividades



acadêmicas, que ora estão à deriva da pesquisa, ora a reboque, se materializando em: disciplinas, cursos livres, linhas de pesquisa, projetos de pesquisa, grupos de estudos, atividades e projetos de extensão, participação em bancas, orientações, revisão de periódicos, organização de eventos, produção técnica, ou ainda, produção científica nas demais tipologias documentais.

Figura 1 - Esquematização das atividades acadêmicas que incluem: publicações, auxílios e outras atividades.



Fonte: Dados da pesquisa (2022).

#### 4.1 Coleta e processamento de dados

A coleta dos dados nas fontes (Web of Science, Biblioteca Virtual FAPESP e Diretório dos grupos de pesquisa no Brasil-DGP/CNPq) foi realizada no mês de janeiro de 2022, utilizando os seguintes termos de busca: “big data” or “humanidades digitais” or “digital humanities” or “humanitas digitalis” or “humanidades computacionais” or “computing humanities”.

Não foi feita qualquer restrição de ano de publicação, assim como não se restringiu os tipos de documentos, para então, serem selecionados os registros com pelo menos um autor de cada uma das universidades estaduais paulistas (UEPs): Universidade de São Paulo (USP), Universidade Estadual de Campinas (UNICAMP) e/ou Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP). Por fim, realizou-se uma padronização dos nomes dos autores e de instituições para se evitar duplicidades.

## **Etapa 1 - Produção científica na Web of Science**

O mapeamento da produção na Web of Science foi realizado na Coleção Principal no campo tópico que engloba: Título, Resumo, Palavras-chave de autor, Keywords Plus®, tendo como resultado 30.268 registros, dos quais 348 tinham autores brasileiros. Após a análise da temática realizada por meio da leitura dos resumos dos documentos recuperados, verificação de duplicidade dos documentos, e restrição a documentos publicados por autores de alguma das UEPs, resultaram 281 registros, com 531 pesquisadores destas universidades. Dentre os documentos, 265 correspondiam a artigos (originais, de revisão ou trabalhos completos em eventos) e os 16 restantes outros documentos (editorias, cartas, resumos, etc.).

## **Etapa 2 - Atividades acadêmicas: Biblioteca Virtual da FAPESP e Diretório dos grupos de pesquisa do CNPq**

### *Biblioteca Virtual FAPESP (BV-FAPESP)*

A pesquisa na BV-FAPESP resultou da utilização dos mesmos termos, na busca avançada, nos campos: Título, Resumo e Assunto, recuperando 131 registros de auxílios. Os filtros dos resultados foram utilizados para selecionar aqueles sob a responsabilidade de pesquisadores das UEPs, resultando em 116 pesquisadores. A análise da temática do auxílio foi realizada por meio da leitura dos resumos dos projetos aprovados, resultando em 94 pesquisadores (que posteriormente passam a ser 93, dada a ausência de currículo Lattes do pesquisador Nicholas Spyridon Vonortas<sup>5</sup>).

### *Diretório dos grupos de pesquisa no Brasil – Lattes (DGP/CNPq)*

No DGP/CNPq foram utilizados os termos para identificação das temáticas de interesse, nos seguintes campos de busca: Grupo, Linha de Pesquisa e Palavras-Chave, com situação certificado. Os filtros região (Sudeste) e Instituição (UEPs) foram utilizados, resultando em 21 grupos de pesquisa com 34 pesquisadores responsáveis<sup>6</sup>. Pelo fato dos

---

<sup>5</sup> Naquele momento era professor na Universidade George Washington *University*, tendo lecionado na UNICAMP entre 2014-2019.

<sup>6</sup> Deve-se considerar que cada grupo pode ter um ou dois líderes, podendo um pesquisador estar presente em mais de um grupo de pesquisa. Os grupos foram mapeados no ano de 2018, tendo sido atualizados em 2022 (resultando no acréscimo de 8 pesquisadores e a não permanência de 9 pesquisadores, cujos grupos não estavam mais ativos, mas foram mantidos pelo fato de apresentarem atividades acadêmicas em seu currículo).



grupos nem sempre apresentarem resumo descritivo, procedeu-se a análise dos currículos Lattes dos pesquisadores para confirmar a aderência à temática, resultando em 29 pesquisadores.

Então realizou-se uma padronização dos nomes de 114 pesquisadores mapeados na BV-FAPESP e DGP/CNPq, observando-se que: 7 deles apresentavam-se em ambas, 21 foram identificados somente no DGP/CNPq e 86 da BV-FAPESP.

### **Etapa 3 - Outras atividades acadêmicas: Plataforma Lattes**

Da Plataforma Lattes, foram exportados os 114 currículos no formato XML, procedendo-se a identificação das seguintes modalidades de atividade acadêmica: atuação profissional (disciplinas ministradas, linhas de pesquisa e projeto de pesquisa); pós-doutorado ou livre-docência; produção bibliográfica (artigos de periódicos, capítulos de livros, livros e trabalhos publicados em anais de eventos); produção técnica (apresentação de trabalho, organização de eventos); orientações, e; participação em bancas. A consideração e contagem das atividades acadêmicas resultaram da análise temática dos títulos dessas modalidades, conforme registro no currículo. A instituição do autor corresponde ao vínculo atual declarado no currículo, apesar de que três deles assinaram artigos com duas UEPs diferentes, o que foi considerado na análise dos artigos recuperados da base Web of Science somente. Um total de 72 currículos continham alguma atividade acadêmica com menção às temáticas Big Data e Humanidades Digitais.

## **5 RESULTADOS**

O estudo considerou um total de 415 pesquisadores das UEPs, sendo que apenas 26 (6,0%) foram localizados nas agências de fomento e publicaram na WoS, 301 (72,5%) somente publicaram na WoS, e 89 (21,4%) constavam das agências, mas não publicaram na WoS.

Esses diferentes grupos de pesquisadores – que praticamente em sua totalidade deve estar em regime de dedicação exclusiva à docência e pesquisa – dão uma ideia de que, apesar da maioria haver publicado na WoS, uma parte significativa não tem publicações indexadas nessa base, mas vem desenvolvendo atividades acadêmicas diversas. Tais atividades, além de publicações em canais diferentes (livros, ou periódicos ou anais de eventos não indexados na WoS), podem advir de interesse de outros indivíduos que convidam os pesquisadores (para bancas de defesa, orientação de alunos, composição



de mesa redonda ou afins), não necessariamente por sua especialização prévia nas temáticas em questão.

## 5.1 Produção científica

A produção científica das UEPs sobre HDs e BD na WoS é representada por 281 documentos (Figura 2), sendo a maior parte destes (93,2%) sobre Big Data. A USP assina o maior percentual de documentos (63,7%), seguida da UNICAMP (26,7%) e UNESP (14,9%).

As temáticas se mostram recentes, dentro do período de 2013 a 2022, com maior quantidade de documentos nos anos de 2018, 2019, 2021 e 2020 (em ordem decrescente do número de documentos). Predominam os artigos originais em periódicos (59,8%), ainda que parte significativa seja publicada em anais de eventos (24,9%), e 9,6,2% de artigos de revisão.

O alto número de publicações em inglês (89,7%) é esperado considerando-se a utilização da WoS na primeira etapa do estudo, assim como a predominância de documentos das Ciências Exatas e da Terra (um montante de 43,8,2% em Ciência da Computação e Engenharia). Conforme apontaram Leite, Mugnaini e Leta (2011) as Ciências Exatas e da Terra são as que apresentam maior percentual de pesquisadores cuja tendência de publicação se dá em outros idiomas (principalmente em inglês). Já os 26 documentos em português são publicados majoritariamente em periódicos do Brasil (16), três (3) em anais de eventos publicados por editora norte americana e os outros dois (2) publicados em periódicos da Colômbia e Espanha.

Figura 2 – Publicações recuperadas na Web of Science, segundo temáticas, período, idioma, área, instituições de afiliação e tipologia documental.

### 281 publicações recuperadas na WoS:

a USP está presente em 63,7%, a UNICAMP em 26,7% e a UNESP em 14,9%;  
61 países estão representados (incluindo o Brasil).

Temáticas	Período	Idioma	Área do conhecimento	Tipologia documental
92,9% - Big Data 6,8% - HDs 0,4% - HDs e BD	2013 até 2022, com destaque para 50 documentos em 2018	89,7% - Inglês 9,3% - Português Outros: 1,1%	36,6% - Ciência da Computação 8,7% - Ciências da Saúde 7,2% - Engenharia	59,8% - Artigos 24,9% - Anais de eventos 9,6% - Revisão



Nota: Dados da pesquisa (2022).

Os Estados Unidos figuram como local de publicação de 37,7% do total de documentos (dados não apresentados), sendo que esse percentual varia significativamente ao se considerar somente artigos originais em periódicos (23,8%), anais de eventos (72,9%) e artigos de revisão (37,0%). Percebe-se a centralidade do país norte americano quando se considera a publicação de anais de eventos, contudo uma maior diversificação é observada entre os países de publicação dos periódicos (artigos originais), que apresenta nas posições seguintes: Inglaterra com 23,6%, Brasil com 14,9%, e Países Baixos e Suíça com 11,8% e 11,3%, respectivamente – restando dez países com percentuais variando de 2,6% a 0,5%.

Tais resultados indicam a necessidade de se compreender a relação entre temática, área do conhecimento e tipologia documental, pois como se observa na Tabela 1 a Ciência da Computação e Engenharia são as principais responsáveis pelas publicações em anais de eventos. Além disso, como se apresentou anteriormente, os eventos ocorreram majoritariamente nos Estados Unidos.

A Ciência da Informação e Biblioteconomia é a área que publica maior percentual de artigos em HDs (38,9%). Já os artigos de revisão estão distribuídos entre as diversas áreas, sendo a Ciência da Informação a única área que apresenta revisão em cada uma das temáticas, e também o único artigo original que trata de ambas as temáticas concomitantemente.

Tabela 1 – Áreas do conhecimento, temáticas das publicações e tipos de documentos (WoS)

Áreas do conhecimento	Tema / Tipo de documento							Total	
	BD			HDs			BD / HDs		
	Artigo	Artigo em anais	Revisão	Artigo	Artigo em anais	Revisão	Artigo		
Ciência da Computação	34	56	2	3	2			97	
Ciências da Saúde	17		6					23	
Engenharia	9	8	2					19	
Ciência da Informação e Biblioteconomia	10		1	5		1	1	18	
Ciências Ambientais e Ecologia	12	1	4					17	
Gestão e Economia	13	2						15	
Química	7		5					12	
Agricultura	8		2					10	
Educação	5		1	1				7	
Outras áreas	38	1	3	5				47	
<b>Total</b>	<b>153</b>	<b>68</b>	<b>26</b>	<b>14</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>265</b>	

Nota: Dados da pesquisa (2022).



As principais fontes de publicações de estudos sobre as temáticas podem ser observadas na Tabela 2, que considera as três áreas que mais publicaram documentos, considerando fontes com ao menos duas publicações. Como se pode notar a dispersão em diversas fontes é evidente, uma vez que dos 97 documentos da área de Ciência da Computação (Tabela 1), apenas 27 constam na Tabela 2, representando menos de um terço dos documentos, enquanto os 73 documentos restantes foram publicados em diferentes fontes de publicação. Soma-se ainda o fato de que apenas 8 das fontes inclui “Big Data” em seu título, sendo 4 deles os *IEEE International Congress on Big Data* (edições de 2013, 2014, 2015 e 2017, sendo que os dois últimos não constam da Tabela 2) e mais três outros eventos e um periódico. Evidencia-se assim que, possivelmente o início de uma especialidade desta área começa a se consolidar neste evento da IEEE (*Institute of Electrical and Electronic Engineers*), ou a comunidade científica das UEPs vem se inserindo principalmente por esta via.

E as áreas de Ciência da Saúde publicaram pelo menos dois documentos, em seis periódicos, podendo-se notar que a área tem direcionado sua pesquisa na temática BD. Estudos revelam que a área da Medicina é um ambiente ideal para a utilização de *Big Data Analysis* (BDA) e *Machine Learning* (ML), devido à enorme quantidade de informações processadas e armazenadas em formato eletrônico, e que essas ferramentas computacionais podem melhorar a pesquisa clínica e tomada de decisões na área de saúde (ORGANIZING COMMITTEE OF THE MADRID 2017 CRITICAL CARE DATATHON *et al.* 2019; NÚNEZ REIZ; ARMENGOL DE LA HOZ; SÁNCHEZ GARCÍA, 2019).

Na Ciência da Informação cinco periódicos se destacaram, sendo quatro deles editados no Brasil e um na Espanha. A ausência de documentos em anais de eventos se deve ao fato de os eventos da área não serem costumeiramente indexados na WoS. Os quatro documentos restantes foram publicados em quatro periódicos diferentes (sendo um brasileiro, três estrangeiros), não constando da Tabela 2.

Por fim, observa-se que a área de Engenharia, apesar de ser a que apresentou o terceiro maior número de documentos, teve apenas uma fonte com duas publicações, não figurando na Tabela 2.

Para uma visualização do ponto de vista da autoria dos artigos, a análise a seguir recorre ao conjunto de 319 autores que publicaram na WoS. A USP apresenta o maior número de autores (186), seguida da UNICAMP com 93 autores, e a UNESP com 42. É importante salientar que dois autores assinaram alguns de seus artigos com dupla



instituição (USP e UNESP), além de haver 6 autores que publicaram em cada uma das temáticas, fazendo com que a soma dessas categorias ultrapasse o total de 326 autores.

Tabela 2 – Áreas do conhecimento (com ao menos 5 documentos), fontes de publicação (com ao menos 2 documentos) e temáticas das publicações (WoS)

Áreas do conhecimento	Tema			Total
	BD	HDs	BD / HDs	
Periódicos / Anais de eventos				
<b>Ciência da Computação</b>				
IEEE ACCESS	2		1	3
EXPERT SYSTEMS WITH APPLICATIONS	3			3
FUTURE GENERATION COMPUTER SYSTEMS-THE INTERNATIONAL JOURNAL OF ESCIENCE	3			3
TECHNOLOGICAL INNOVATION FOR INDUSTRY AND SERVICE SYSTEMS, DOCEIS 2019	2			2
JOURNAL OF NETWORK AND COMPUTER APPLICATIONS	2			2
2014 IEEE WORLD CONGRESS ON SERVICES (SERVICES)	2			2
ACM COMPUTING SURVEYS	2			2
PATTERN RECOGNITION LETTERS	2			2
2013 IEEE INTERNATIONAL CONGRESS ON BIG DATA	2			2
2014 IEEE INTERNATIONAL CONGRESS ON BIG DATA (BIGDATA CONGRESS)	2			2
2014 IEEE 10TH INTERNATIONAL CONF. ON ESCIENCE WORKSHOPS (ESCIENCE 2014)	1		1	2
IEEE TRANSACTIONS ON IMAGE PROCESSING	1		1	2
<b>Ciências da Saúde</b>				
CLINICAL AND EXPERIMENTAL RHEUMATOLOGY	5			5
ANNALS OF THE RHEUMATIC DISEASES	4			4
RHEUMATOLOGY	4			4
NEUROIMAGE	2			2
ARTHRITIS & RHEUMATOLOGY	2			2
EINSTEIN-SAO PAULO	2			2
<b>Ciência da Informação e Biblioteconomia</b>				
ENCONTROS BIBLI-REVISTA ELETRONICA DE BIBLIOTECONOMIA E CIENCIA DA INFORMACAO	3		1	4
REVISTA IBERO-AMERICANA DE CIENCIA DA INFORMACAO		1	2	3
EM QUESTAO	2		1	3
INTERNATIONAL JOURNAL OF INFORMATION MANAGEMENT	2			2
IBERSID-REVISTA DE SISTEMAS DE INFORMACION Y DOCUMENTACION	1		1	2
<b>Total</b>	<b>51</b>	<b>1</b>	<b>8</b>	<b>60</b>

Nota: Dados da pesquisa (2022).

O Gráfico 1 apresenta a distribuição dos pesquisadores segundo os tipos de colaboração institucional – estabelecidos por meio da coautoria nos artigos. As três instituições de origem dos autores (UEPs) são combinadas com demais tipos de instituição (outra UEP, outra brasileira e/ou estrangeira) ou sem colaboração, conforme cores das barras. Pelo fato de os perfis de colaboração institucional terem sido identificados para cada um dos 531 autores das UEPs, somam-se 719 perfis. Nas temáticas, os perfis foram considerados, independente da UEP dos autores, fazendo com que o total de perfis dos 181 artigos chegasse a 319.



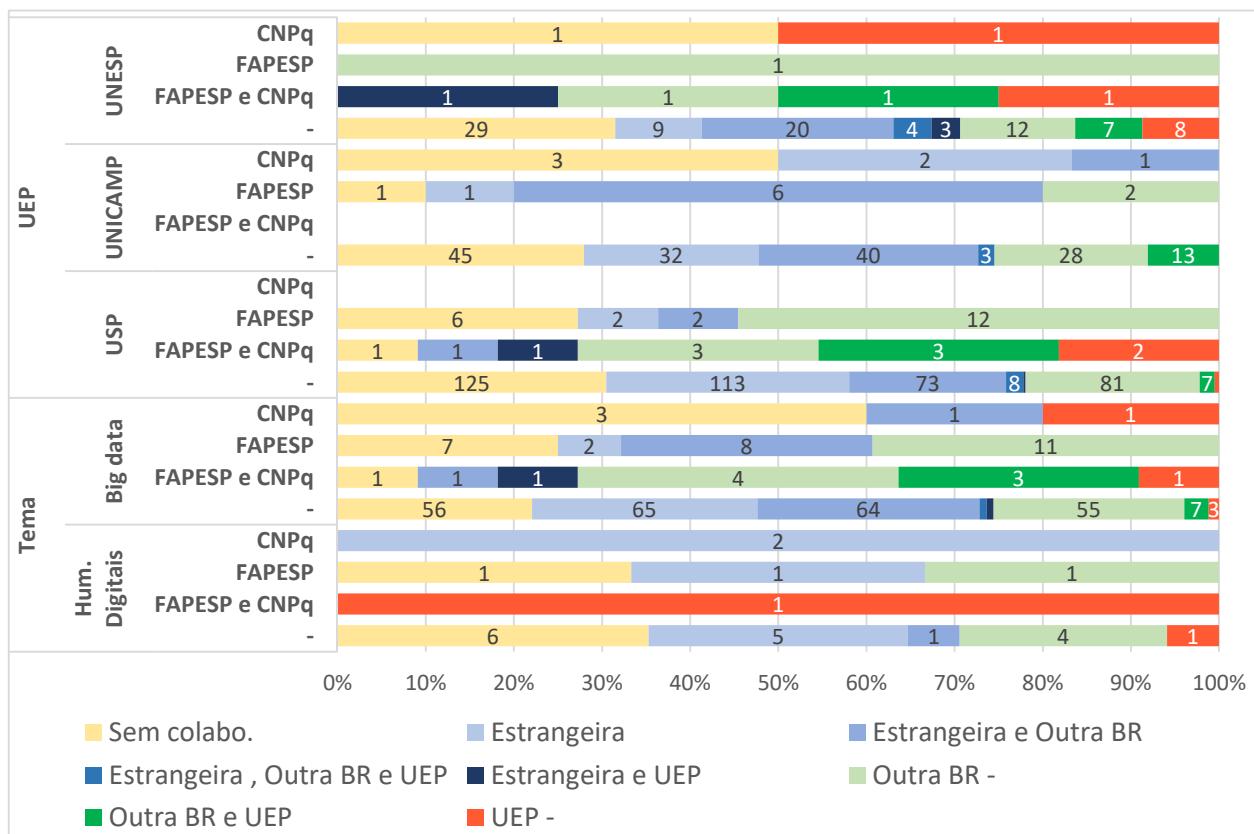
Sobre presença de artigos de pesquisadores vinculados às agências de fomento, nota-se que da USP predominam aqueles com auxílios identificados na BV-FAPESP e alguns que também lideram grupos registrados no DGP/CNPq. Na UNICAMP observam-se alguns pesquisadores que são identificados, na fonte de uma agência ou da outra, mas não em ambas. E da UNESP apenas um artigo de pesquisador vinculado a auxílio FAPESP, dois vinculados a grupos registrados no DGP/CNPq, e quatro a ambas as fontes.

Pode-se observar no Gráfico 1 uma tendência de colaboração com instituições estrangeiras entre autores não mapeados nas agências, o que fica claro devido ao volume observado nessa categoria. Ainda na categoria dos não mapeados, a UNESP é a que apresenta perfil mais diversificado, com todos os tipos de colaboração observados, seguida da mesma categoria na temática BD, USP e UNESP.

A USP apresenta dois autores que receberam auxílio FAPESP e lideram grupo registrado no CNPQ, que são: Fabio Kon, que é da área de Ciências da Computação, publicou exclusivamente em BD e tem 4 artigos em colaboração com outras instituições brasileiras, sendo que um deles inclui dois autores dos Estados Unidos; José Eduardo Santarem Segundo, que também assina seus artigos como UNESP (por estar credenciado em Programa de Pós-Graduação em Ciência da Informação daquela universidade), publicou 6 artigos em BD e um em HDs, sendo um destes em colaboração com pesquisadora canadense, e os demais com pesquisadores brasileiros (da UNESP e outras instituições brasileiras); e Francisco Carlos Paletta, da área de Ciência da Informação, publicou um artigo em autoria única na temática BD. A existência destes autores mapeados nas agências de fomento, somado à colaboração com outra UEP e/ou outras instituições brasileiras, indica execução de atividades acadêmicas, que podem fomentar institucionalização e nucleação nacional.



Gráfico 1 – Distribuição dos pesquisadores das UEPs que publicaram na WoS, segundo instituição, temática, agência de fomento e tipo de colaboração institucional nos documentos.



Fonte: Dados da pesquisa (2022).

Considerando a diversidade de tipos de colaboração, convém apresentar alguns dados não apresentados graficamente, sobre os países das instituições de coautores estrangeiros. Os Estados Unidos figuram como país colaborador mais frequente, participando de 19,6% do total de artigos, 24,6% dos artigos da USP, 12,0% da UNICAMP e 7,1% da UNESP. Apesar da Inglaterra estar na segunda posição, considerando-se o total de artigos das UEPs (com 12,1%), o Canadá é o principal colaborador da UNICAMP (com 13,3% de seus artigos), e a França assina 16,7% dos artigos da UNESP.

Pode-se destacar a China na sétima posição de colaboradores da USP, dividindo posição com a Itália. E ainda Portugal como terceiro país colaborador mais frequente da UNICAMP e Espanha o terceiro mais frequente da UNESP.

A temática BD praticamente determina o ranking do total de artigos, já que representa 92,9% dos artigos, apresentando nas 11 primeiras posições os seguintes países

com pelo menos 19 artigos: Estados Unidos, sete países da Europa Ocidental, Canadá, Austrália e China. E nas HDs, dos 19 artigos, 8 apresentaram colaboradores estrangeiros, sendo 3 dos Estados Unidos, 3 de Portugal, 2 da França e outros 4 países que colaboraram uma única vez. E finalmente, o único artigo que trata de ambas as temáticas foi assinado por duas autoras da USP - Brasilina Passarelli e Ana Cláudia Fernandes Gomes.

## 5.2 Atividades acadêmicas

Nesta seção foram analisadas as atividades acadêmicas dos 114 pesquisadores registrados nas fontes das agências de fomento. Constituíram-se de informações sobre a atividade acadêmica sobre HDs e BD, de pesquisadores das universidades estaduais paulistas (USP, UNICAMP e UNESP) recuperadas: na Biblioteca Virtual da FAPESP (BV-FAPESP) e no Diretório dos grupos de pesquisa no Brasil-Lattes (DGP/CNPq) (auxílios à pesquisa), e; na Plataforma Currículo Lattes.

### 5.2.1 *Projetos de pesquisa financiados pela FAPESP*

A FAPESP apoiou 130 projetos relacionados às temáticas, sendo que em BD a UNESP teve 17,9% (19), UNICAMP 23,6% (25) e USP 58,4% (62). A temática de HDs teve 6 fomentos para UNESP 25% e UNICAMP 25%, e USP com 50% financiamentos (12), perfazendo um total de 24 projetos financiados.

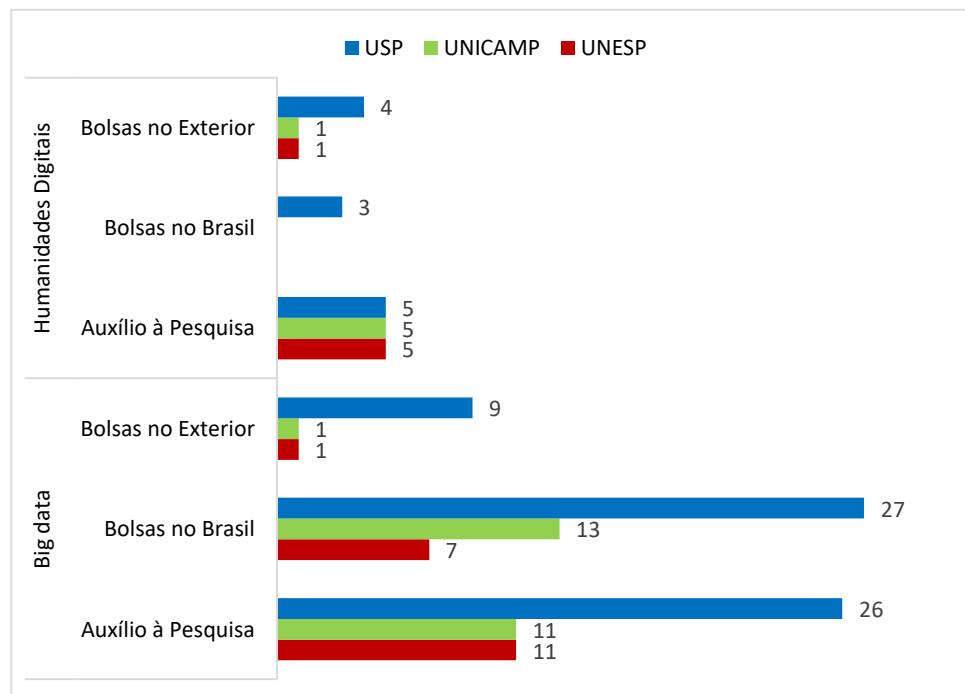
Os auxílios à pesquisa totalizam 63 tipos, conforme Gráfico 2: organização de reunião científica, auxílio regular, pesquisador visitante nacional e internacional, auxílios temáticos, reunião no Brasil e no exterior, apoio a jovens pesquisadores, programa e-science e data science e parceria para inovação tecnológica (PITE). Nessa linha de fomento, a temática BD foi contemplada em 76% (48) dos auxílios, sendo que os pesquisadores da USP receberam 26 auxílios, UNESP e UNICAMP receberam 11 fomentos cada uma. Para ambas as temáticas, destacam-se especialmente os auxílios a organização de reunião científica (BD=15 e HDs=8) e auxílio regular, (BD=11 e HDs=3).

Na linha Bolsas no País destaca-se a UNICAMP com 11 bolsas de Treinamento Técnico relacionada a temática BD, seguida USP com 5 e UNESP com 4. Sobre a temática de Humanidades digitais se destaca a USP com duas bolsas de Doutorado e uma bolsa de



Pós-Doutorado. Na linha de Bolsas de Pesquisa no Exterior, a UNICAMP obteve uma bolsa em cada uma das temáticas, a USP teve 3 bolsas de pesquisa sobre a temática BD e uma bolsa relacionada a HDs.

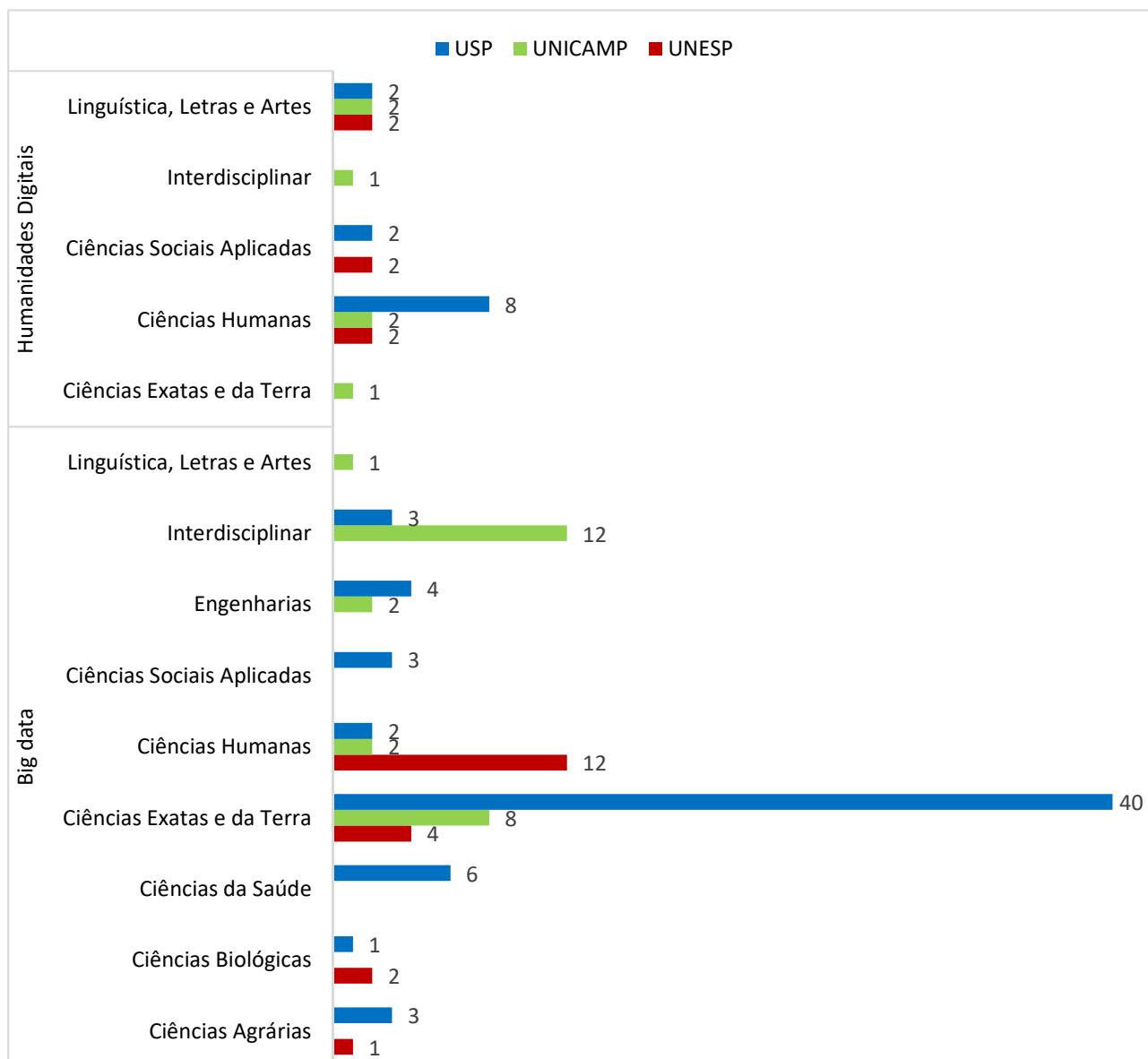
Gráfico 2 – Fomento da FAPESP, segundo tipo, instituição e temática.



Nota: Dados da pesquisa (2022).

No que se refere às áreas do conhecimento (Gráfico 3), a área que mais se destacou da USP foram as Ciências Exatas e da Terra com 64,5% (40) do total de 62 fomentos relacionado a BD. Em relação a temática HDs as Ciências Humanas obtiveram 8 fomentos, Linguística, Letras e Artes tiveram 2 e Ciências Sociais Aplicadas receberam 2 auxílios também. Diferentemente do resultado encontrado na USP na UNICAMP a área do conhecimento em destaque foi a Interdisciplinar com 48% (12) do total de 25 fomentos relacionados à temática BD. Nas HDs totalizou 6 bolsas, sendo que Linguística, Letras e Artes e Ciências Humanas tiveram 2 bolsas em cada área e Ciências Exatas e da Terra e Interdisciplinar 1 bolsa cada uma. Na UNESP a temática do Big Data foi contemplada em 19 projetos, sendo que 12 foram de pesquisadores das Ciências Humanas, 4 das Ciências Exatas e da Terra, 2 das Ciências Biológicas e apenas 1 das Ciências Agrárias. A temática das Humanidades Digitais, assim como na UNICAMP, foi abordada em 6 fomentos, divididos igualmente entre as Ciências Humanas (2), Linguística, Letras e Artes (2) e Ciências Sociais Aplicadas (2).

Gráfico 3 – Temáticas e áreas do conhecimento dos projetos que receberam financiamento da FAPESP, segundo instituição.



Nota: Dados da pesquisa (2022).

Ressalta-se que a área de Ciências Exatas e da Terra da USP se destaca na quantidade de fomentos na temática Big Data, principalmente pelos pesquisadores do Instituto de Matemática e Estatística (IME), Instituto de Ciências Matemáticas e de

Computação (ICMC), Escola Politécnica (EP), Instituto de Física de São Carlos (IFSC), relacionados aplicação no Big Data nas subáreas do conhecimento: Metodologia e Técnicas da Computação, Probabilidade e Estatística Aplicadas e Sistemas de Computação.

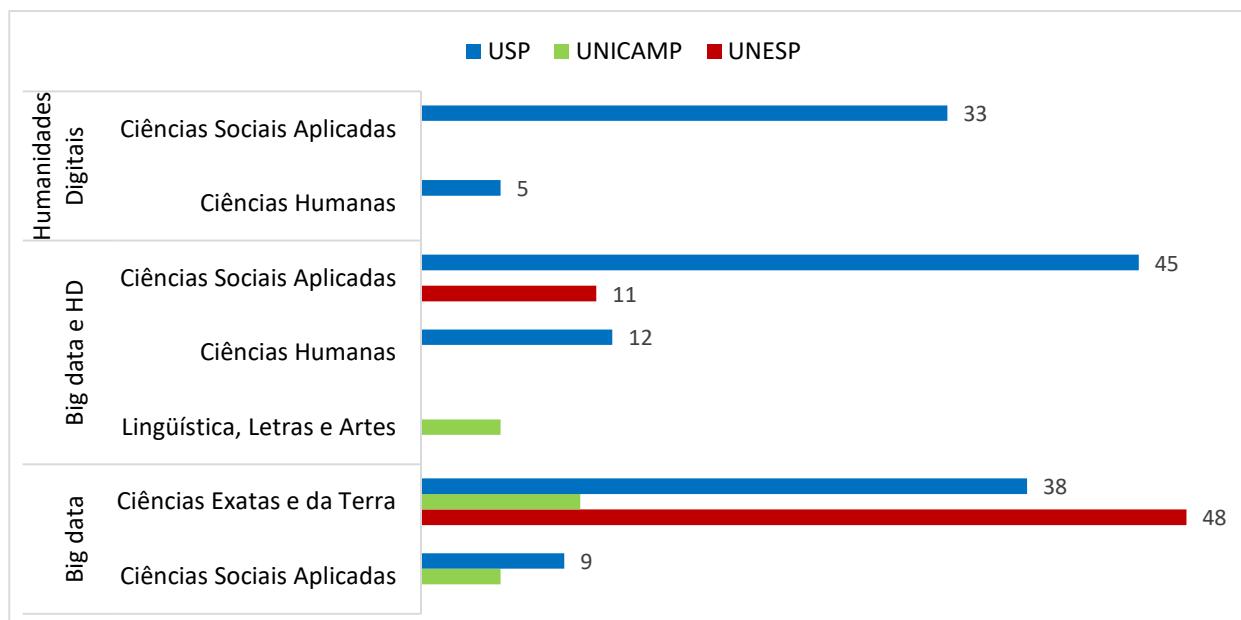
### **5.2.2 Grupos de pesquisa disponíveis no Diretório dos Grupos de Pesquisa do CNPq**

O DGP/CNPq possibilitou a recuperação de informações sobre os grupos de pesquisa certificados, tais como: nomes dos pesquisadores; às linhas de pesquisa; às especialidades do conhecimento; aos setores de aplicação envolvidos, entre outras.

O *corpus* da pesquisa realizada no DGP/CNPq é composto de 29 pesquisadores, sendo 58,6% da USP (17), 24,1% da UNICAMP (7) e 17,2% da UNESP (5). Seus grupos de pesquisa dedicam-se principalmente à temática BD (62% ou 18), ambas as temáticas (28% ou 8), ou HDs (10% ou 3). A área de Ciências Sociais Aplicadas é a que mais se destaca com 103 atividades, sendo que a USP consta em todas as temáticas. A UNICAMP aparece 5 somente na temática BD e UNESP com 11 na temática BD/HDs. Na área de Ciências Exatas e da Terra a única temática que se destaca é BD com 96, sendo UNESP 48, UNICAMP 10 e USP 38. Nas Ciências Humanas a única universidade que se destaca é USP com 12 atividades de BD e 5 de HDs, e na área de Lingüística, Letras e Artes temos 5 atividades atribuídas a pesquisadores da UNICAMP.

Gráfico 4 – Total de atividades dos pesquisadores que receberam fomento do CNPq subdivididos por área, instituição e temáticas.





Nota: Dados da pesquisa (2022).

Pelo fato do estudo não se limitar somente à produção científica, como a maioria dos estudos bibliométricos/cientométricos, foi possível captar o estado da arte atual das temáticas nas agências de fomento e por consequência nas pesquisas realizadas nas UEPs. Isso porque, se as análises se limitassem apenas a análise do primeiro corpus (WoS) representaria principalmente as temáticas em áreas de Ciências Exatas e da Terra. A maior abrangência das fontes adotadas (corpora de auxílios) permitiu observar o papel que as Humanidades e Ciências Sociais Aplicadas vêm desempenhando, e mais importante, a inserção da Ciência da Informação.

### 5.2.3 Atividades acadêmicas nos currículos Lattes

Destaca-se a relevância de se analisar as atividades acadêmicas relacionadas às temáticas disponíveis na Plataforma Lattes dos pesquisadores mapeados nas fontes das agências FAPESP e/ou CNPq, pois o propósito da Plataforma é o de gerenciar os dados sobre profissionais em ciência e tecnologia no Brasil. Assim, o preenchimento adequado e a atualização de acordo com as atividades acadêmicas desenvolvidas são relevantes para dar visibilidade individual e institucional. As atividades acadêmicas de 72 pesquisadores foram identificadas, que se distribuem da seguinte forma: em relação às temáticas, 17,4% mencionam HDs, 19,8% ambas as temáticas e 62,8% BD; e quanto às agências de fomento, 61,1% apresentam algum auxílio FAPESP, 29,2% lideram grupo do DGP/CNPq e 9,7% atende os dois critérios. Percebe-se assim que os 51 pesquisadores com auxílio

FAPESP representam a maioria dos pesquisadores com atividades (70,8%). Por outro lado, os restantes 42 pesquisadores que não apresentam atividades são todos beneficiários de auxílio FAPESP. A ausência de atividades, nesse caso, se deve ao fato de pesquisadores apresentarem auxílios que mencionem as temáticas de forma tangencial (no resumo ou palavras-chave), contudo sem qualquer menção no currículo.

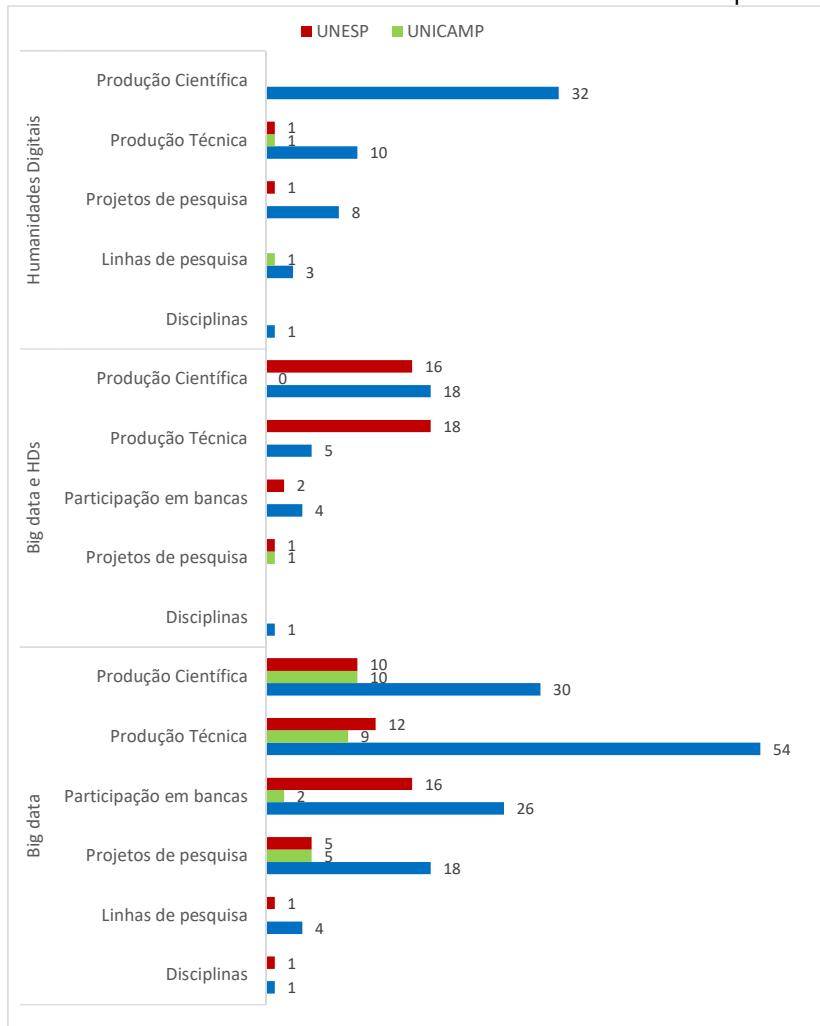
A categoria **produção científica** se destaca pois representa 34,8% (116) do total de atividades relacionadas às temáticas. Os 47 pesquisadores da USP com atividades nas temáticas representam 65,3% do total de pesquisadores nesta etapa da análise (que é de 72), enquanto que seu volume de produção científica equivale a 66,1% do total. A UNICAMP tem 14 pesquisadores (19,4%) e representa 8,7% da produção científica total, enquanto a UNESP, com 11 pesquisadores (ou 15,3%) perfaz 25,2% da produção científica, evidenciando assim maior atividade em relação à UNICAMP. Os trabalhos completos em anais de eventos representam 58,6% das publicações, seguidos dos artigos em periódicos (25,0%), capítulos de livros (13,8%) e livros (2,6%, que consistem de apenas três livros, sendo um em HDs e dois em ambas as temáticas).

No Gráfico 5 pode-se observar que o maior volume de produção científica da UNESP se dá majoritariamente em publicações que tratam sobre as duas temáticas, enquanto que a USP apresenta maior volume de publicação em cada temática, individualmente, e a UNICAMP mostra uma concentração específica em BD.

A **produção técnica** equivale a 33,0% (110) do total, tendo apresentações de trabalho como a atividade mais numerosa (69,6% em BD e 45,8% em HDs). Dentre as demais atividades, as duas mais frequentes em cada temática são: em BD, curso de curta duração (9,8%) e organização de evento (8,7%), e; em HDs, publicação em mídia social, site ou blog (25%) e organização de evento (12,5%). A produção do Lattes diz respeito às produções intelectuais desenvolvidas pelos pesquisadores que englobam a produção bibliográfica e técnica.



Gráfico 5 – Total de atividades acadêmicas relacionadas às temáticas disponíveis nos currículos



Nota: Dados da pesquisa (2022).

As **participação em bancas** de trabalhos de conclusão ou bancas de comissões julgadoras representam 15% das atividades, relacionadas majoritariamente a BD, e reunindo principalmente pesquisadores de USP e UNICAMP. E os **projetos de pesquisa** representam 15,0%, concentrando-se principalmente em BD (28, sendo 18 na USP) e HDs (9, sendo 8 na USP), e apenas dois em ambas temáticas (um da UNICAMP e outro da UNESP).

Na categoria atuação profissional reúnem-se as atividades de ensino (**disciplinas** em Graduação e Pós-Graduação) e **linhas de pesquisa**, as quais são representadas por “temas aglutinadores de estudos científicos que se fundamentam em tradição investigativa, de onde se originam projetos cujos resultados guardam afinidades entre si” (PLATAFORMA LATTES, [2022]). Quatro disciplinas de pós-graduação foram identificadas, sendo 3 na USP: uma em BD, oferecida por Pedro Luiz Pizzigatti Corrêa, da Ciência da Computação;

outra em HDs, de Maria Clara Paixão de Sousa, de Letras; e a de Francisco Carlos Paletta, supracitado, em ambas as temáticas) e uma na UNESP em BD. E da UNESP, José Marques Júnior, da Agronomia.

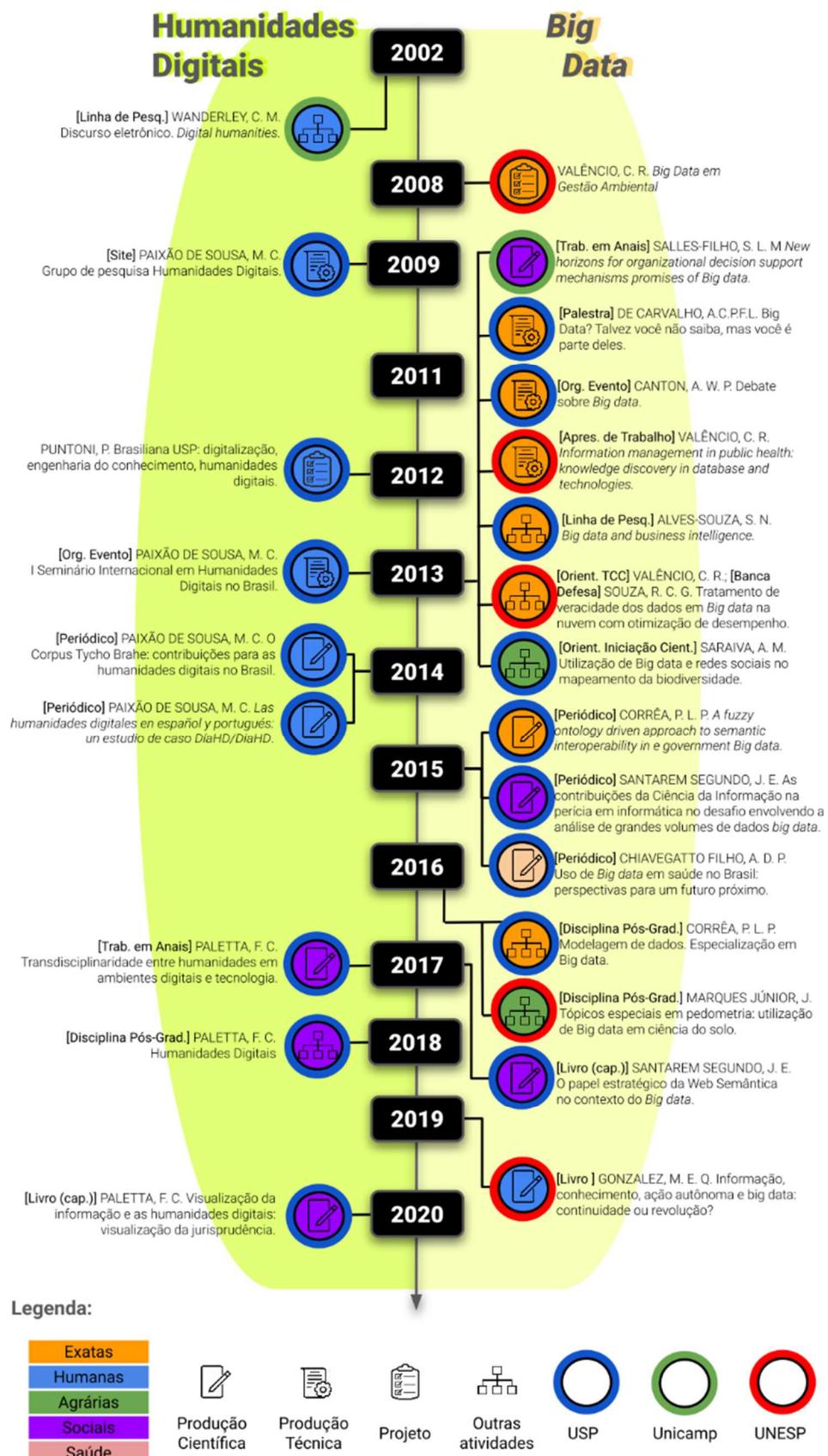
Para se ter uma visão da evolução das atividades em cada temática, foram identificados os anos de registro da primeira ocorrência de cada um dos tipos de atividades nos currículos dos pesquisadores. O infográfico a seguir (Figura 3, assim como um Apêndice) permite observar que as primeiras atividades em HDs são exclusivas das áreas de Ciências Humanas e Sociais Aplicadas. Além disso, concordando com Hu e Zhang (2017), ocorrem antes de BD, estreando em 2002 no currículo de um pesquisador da UNICAMP, que registra a Linha de Pesquisa na Filosofia.

Todas as atividades em HDs são de pesquisadores da USP, sendo a primeira delas em 2009, por um grupo de pesquisa denominado Humanidades Digitais, que tem seu site registrado na área de Letras. O mesmo grupo organiza em 2013 o já mencionado 1º Seminário Internacional em Humanidades Digitais no Brasil, e em autoria de sua líder Maria Clara Paixão de Sousa', tem os primeiros artigos em periódico publicados (2014).

O primeiro projeto de pesquisa em HDs foi registrado em 2012 na área de História, e o primeiro trabalho em anais de evento ocorre em 2017, de Francisco Carlos Paletta (da Ciência da Informação), que oferece uma disciplina de pós-graduação no ano seguinte e publica um capítulo de livro em 2020.



Figura 3 – Primeiras atividades de cada tipo disponíveis nos currículos, segundo temática, ano, área e UEP



Nota: Dados da pesquisa (2022).

Na temática de BD a primeira atividade é um projeto de pesquisa de 2008, registrado por um pesquisador da UNESP da área de Ciência da Computação, aplicando BD em gestão ambiental. As atividades subsequentes se dão somente em 2013, que é o ano com maior ocorrência nos currículos, totalizando sete atividades, sendo: quatro na USP, duas na UNESP e uma na UNICAMP. O primeiro trabalho em anais se deu na área de Economia, enquanto que a primeira orientação foi de uma iniciação científica, em Ciências Agrárias. As outras cinco atividades são da área de computação, sendo três produções técnicas (palestra, organização de evento e apresentação de trabalho), uma linha de pesquisa e uma orientação de TCC (trabalho de conclusão de curso, que também é registrado por uma pesquisadora que participou da banca de defesa).

Já no ano de 2015 se dão os primeiros artigos em periódico, das áreas de Ciência da Computação, Ciência da Informação e Saúde Coletiva. Em 2016 surgem as duas disciplinas de pós-graduação, supracitadas (da USP e UNESP). E finalmente, observam-se as publicações mais recentes, um capítulo de livro em 2017 (da USP, Ciência da Informação), e o único livro – de 2019, que consiste de uma coletânea organizada por três docentes da UNESP, sendo Maria Eunice Quilici Gonzalez (da Filosofia) a pesquisadora componente desta amostra por apresentar 9 auxílios à pesquisa na BV-FAPESP.

Como se pode perceber o início do período, até 2013, é marcado pelo surgimento de uma diversidade de atividades acadêmicas e produções técnicas, enquanto que no período restante predominam as produções científicas, marcadas pelos diferentes canais de publicação. Ainda que de maneira geral – por se tratar apenas dos primeiros registros de cada tipo de atividade, desconsiderando o volume de cada uma, ou as demais atividades desenvolvidas pelos pesquisadores subsequentemente – são esboçadas trilhas de desenvolvimento da pesquisa, em cada área, segundo as características específicas de cada temática.

## 6 CONSIDERAÇÕES FINAIS

Esta pesquisa evidenciou que a opção por uma abordagem mais ampla de análise das atividades acadêmicas – incluindo tanto as pesquisas realizadas (produções) como auxílios à pesquisa (projetos e bolsas, que são pesquisa potencial), e ainda, outras atividades acadêmicas – permite uma análise mais detalhada da adoção das temáticas de



Humanidades Digitais e Big Data como objeto de interesse dos pesquisadores das universidades estaduais paulistas.

Como se pôde perceber, apesar de uma maior quantidade de pesquisadores das UEPs que colaboraram para a produção científica presente na Web of Science, uma pequena parte apresentava projetos ou bolsas financiadas pela FAPESP e/ou figuravam como líderes de grupos de pesquisa do DGP/CNPq. Por outro lado, cerca de um quinto dos pesquisadores foram mapeados nas agências de fomento, contudo não publicaram na WoS. Este cenário justificou a importância de se conhecer os tipos de atividades desenvolvidas nas temáticas em questão.

A produção indexada na WoS é majoritariamente relacionada à temática de Big Data (cerca de 93%), tendo a Ciência da Computação, Ciências da Saúde e Engenharias como principais áreas de conhecimento, e totalizando pouco mais que a metade do total. A Ciência da Informação veio em seguida, sendo a área mais representativa das Ciências Sociais Aplicadas, e a que mais publicou em Humanidades Digitais. Por conta da indexação de periódicos do Brasil de Ciência da Informação na WoS, quatro periódicos se destacaram entre os mais utilizados desta área. Tal fato é a principal exceção desta produção, cujas características predominantes refletem a ciência *mainstream* (abundantemente de ciências naturais, no idioma inglês, e em periódicos editados nos Estados Unidos e Inglaterra) que configura a WoS. A colaboração internacional se deu principalmente com autores estadunidenses e ingleses (nos dados gerais), contudo os segundos países colaboradores mais frequentes da UNICAMP e UNESP foram, respectivamente, Canadá e França.

Dentre as atividades acadêmicas, os auxílios financiados pela FAPESP, que consistiram de projetos e bolsas no Brasil e no exterior, apresentam um montante mais expressivo (cerca de 18%) relacionado à temática Humanidades Digitais. De modo geral, pôde-se observar uma maior presença das Ciências Humanas, Sociais Aplicadas e Linguística Letras e Artes, em relação aos dados da WoS. Contudo, chamou a atenção o número de auxílios que a UNESP apresentou em Ciências Humanas, na temática Big Data.

Já entre os grupos de pesquisa cadastrados no DGP/CNPq, as Humanidades Digitais ganharam mais atenção do que nas fontes anteriores, pelo fato de 10% se dedicarem exclusivamente a esta temática, e outros 21% identificarem ambas as temáticas. E quando se considerou o volume de atividades declaradas nos currículos dos líderes destes grupos, as Ciências Sociais Aplicadas são as que mais se destacaram, principalmente quando relacionado às Humanidades Digitais. Exclusivamente em Big Data,



as atividades se dão nas Ciências Exatas e da Terra, tendo UNESP e USP como protagonistas.

Finalmente, as atividades acadêmicas declaradas nos currículos de 72 pesquisadores evidenciam a maior dedicação exclusivamente ao Big Data (75%), contudo o volume de atividades destaca ainda mais as Humanidades Digitais em relação às fontes anteriores, com 17,4% exclusivamente e outros 19,8% em ambas as temáticas.

Entender o atual estado das atividades acadêmicas sobre as temáticas nas universidades estaduais paulistas e os esforços realizados em prol do desenvolvimento científico nos permite conhecer as características centrais da informação digital, contribuindo assim não somente para questões pontuais que permeiam os processos acadêmicos e profissionais, mas também auxiliar o desenvolvimento social, econômico e cultural, visto que a informação disposta em formato digital atribui novos significados para a sociedade.

## REFERÊNCIAS

- BARACHO, Renata Maria Abrantes et al. Ciência da informação: sinalizações para o presente e futuro. **XVI Encontro Nacional de Pesquisa em Ciência da Informação**, João Pessoa, v. 16, 2015. Anais de eventos. Disponível em: <http://www.brappci.inf.br/index.php/article/download/43976>. Acesso em: 15 jul. 2018.
- BARRETO, Aldo de Albuquerque. A aventura de perceber significados. **DataGramZero - Revista de Ciência da Informação**, João Pessoa, v. 15, n. 3, p. Artigo 01, 2014. Disponível em: [http://www.academia.edu/7541212/A\\_aventura\\_de\\_perceber\\_significados](http://www.academia.edu/7541212/A_aventura_de_perceber_significados). Acesso em: 28 maio 2018.
- CALDAS, Max Silva; SILVA, Emanoel Costa Claudino. Fundamentos e aplicação do Big Data: como tratar informações em uma sociedade de yottabytes. **Bibl. Univ. - Bibliotecas Universitárias: pesquisas, experiências e perspectivas**, Belo Horizonte, v. 3, n. 1, p. 65-83, jan./jun, 2016. Disponível em: <http://www.brappci.inf.br/index.php/res/v/110347>. Acesso em 13 Jul. 2018.
- CHEN, Michelle. Emerging research trends in big data: big data and digital humanities. **SJSU School Of Information**. Washington, p. 1-3. nov. 2015. Disponível em: <https://ischool.sjsu.edu/ciri-blog/emerging-research-trends-big-data-big-data-and-digital-humanities>. Acesso em: 30 set. 2022.
- CUNNINGHAM, Leigh. The librarian as digital humanist: the collaborative role of the research library in Digital Humanities Projects. **Faculty of Information Quarterly**, Toronto, v. 2, n. 2, Feb./Mar. 2010. Disponivel em: <https://hdl.handle.net/1807/80171> Acesso em: 30 jun. 2022.
- DIGITAL RESEARCH INFRASTRUCTURE FOR THE ARTS AND HUMANITIES (DARIAH); COMMON LANGUAGE RESOURCES AND TECHNOLOGY INFRASTRUCTURE (CLARIN). **Digital Humanities course registry**. [2022]. Disponível em: <https://dhcr.clarin-dariah.eu/>. Acesso em: 22 abril 2019.



DAVYT, Amilcar; VELHO, Léa. A avaliação da ciência e a revisão por pares: passado e presente. Como será o futuro?. *História, Ciências, Saúde-Manguinhos*, vol.7, n.1, p.93-116, 2000. Disponível em: <<https://doi.org/10.1590/S0104-59702000000200005>>. Acesso em 23 ago. 2018.

DUTRA, Moises Lima; MATIAS, Marcio R. Visualização de correspondências semânticas no universo Big Data. **XV Encontro Nacional de Pesquisa em Ciência da Informação**, Belo Horizonte, v. 15, p. 4250-4257, 2014. Anais de eventos. Disponível em: <http://enancib2014.eci.ufmg.br/documentos/anais/anais-gt8>. Acesso em 31 mai. 2018.

FAGUNDES, Priscila Basto; MACEDO, Douglas Dylon Jeronimo; FREUND, Gislaine Parra. A produção científica sobre qualidade de dados em Big Data: um estudo na base de dados Web of Science. **RDBCI - Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v.16, n.1, p. 194-210, jan./abr. 2018. Disponível em: <https://doi.org/10.20396/rdbc.v16i1.8650412>. Acesso em 11 jul. 2018.

FITZPATRICK, Kathleen. The Humanities, Done Digitally. In: GOLD, Matthew K. **Debates in the Digital Humanities**. U de Minnesota Press, 2012. Disponível em: <http://dhdebates.gc.cuny.edu/debates/text/30>. Acesso em: 26 set. 2018.

HU, Jiming; ZHANG, Yin. Discovering the interdisciplinary nature of Big Data research through social network analysis and visualization. **Scientometrics**, Budapest, v. 112, n. 1, p.91-109, 2017. DOI 10.1007/s11192-017-2383-1

JONES, Steven E. The Emergence of the Digital Humanities (as the network is exerting). In: GOLD, Matthew K. **Debates in the Digital Humanities**. U de Minnesota Press, 2012. Disponível em: <http://dhdebates.gc.cuny.edu/debates/text/52>. Acesso em: 09 out. 2018.

KALADHAR, Arnepalli; DORASWAMY NAICK, B. R.; SOMASEKHARA RAO, K. Application of Big Data Technology to Library data: a review. **International Journal of Library and Information Studies**. Vol.8 (2), p. 25-35, Apr-Jun, 2018. Disponível em: [http://www.ijlis.org/img/2018\\_Vol\\_8\\_Issue\\_2/25-30.pdf](http://www.ijlis.org/img/2018_Vol_8_Issue_2/25-30.pdf). Acesso em: 13 set. 2018.

KAPLAN, Frederic. A map for big data research in digital humanities. **Frontiers in Digital Humanities**, 06 de maio de 2015. Disponível em: <http://dx.doi.org/10.3389/fdigh.2015.00001> Acesso em: 30 set. 2022.

KIRSCHENBAUM, Matthew. What is Digital Humanities and what's it doing in English Departments?. In: GOLD, Matthew K. **Debates in the Digital Humanities**. U de Minnesota Press, 2012. Disponível em: <http://dhdebates.gc.cuny.edu/debates/text/38>. Acesso em: 26 set. 2018.

LEITE, Paula; MUGNAINI, Rogério; LETA, Jacqueline. A new indicator for international visibility: exploring Brazilian scientific community. **Scientometrics**, v. 88, p. 311-319, 2011.

LIMA, José Leonardo Oliveira; ALVARES, Lillian. Organização e representação da informação e do conhecimento. In: ALVARES, Lillian (Org.). **Organização e representação da informação e do conhecimento**: conceitos, subsídios interdisciplinares e aplicações. São Paulo: B4 Ed, 2012. p. 21-47.

LIMA JUNIOR, Walter Teixeira. Jornalismo computacional em função da "Era do Big Data". **LÍBERO - Revista do Programa de Pós-Graduação da Faculdade Cásper Líbero**, São Paulo, v. 14, n. 28, p. 45-52, dez. 2011. Disponível em: <http://seer.casperlibero.edu.br/index.php/libero/article/view/329/303>. Acesso em: 15 maio 2018.



MAYER-SCHONBERGER, Viktor. **Big Data**: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana. Tradução de Paulo Polzonoff Junior. 1 ed. Rio de Janeiro: Elsevier, 2013.

MOSTAFA, Solange Puntel; CRUZ, Denise Viuniski N.; AMORIM, Igor Soares. Primavera nos dentes: fuga e resistência na era digital. **Liinc em revista - Laboratório Interdisciplinar sobre Informação e Conhecimento**, Rio de Janeiro, v.11, n.2, p. 360-374, novembro 2015. Disponível em: <http://revista.ibict.br/liinc/article/view/3665/3113>. Acesso em: 31 maio 2018.

MOURA, Maria Aparecida. Ciência da Informação e humanidades digitais: mediações, agência e compartilhamento de saberes. **Perspectivas em Ciência da Informação**, v. 24, n. spe, p. 57-69, 2019. Disponível em: <https://doi.org/10.1590/1981-5344/3893>

NÚÑEZ REIZ, A.; ARMENGOL DE LA HOZ, M. A.; SÁNCHEZ GARCÍA, M. Big data analysis and machine learning in intensive care units. **big data analysis y machine learning en medicina intensiva**. **Medicina intensiva**, v.43, n.7, 416–426., 2019. DOI: <https://doi.org/10.1016/j.medin.2018.10.007>

ORGANIZING COMMITTEE OF THE MADRID 2017 CRITICAL CARE DATATHON, NÚÑEZ REIZ, A., MARTÍNEZ SAGASTI, F., ÁLVAREZ GONZÁLEZ, M., BLESA MALPICA, A., MARTÍN BENÍTEZ, J. C., NIETO CABRERA, M., DEL PINO RAMÍREZ, Á., GIL PERDOMO, J. M., PRADA ALONSO, J., CELI, L. A., ARMENGOL DE LA HOZ, M. Á., DELIBERATO, R., PAIK, K., POLLARD, T., RAFFA, J., TORRES, F., MAYOL, J., CHAFER, J., GONZÁLEZ FERRER, A., ... SÁNCHEZ GARCÍA, M. Big data and machine learning in critical care: opportunities for collaborative research. **Medicina intensiva**, v.43, n.1, p. 52–57, 2019. DOI: <https://doi.org/10.1016/j.medin.2018.06.002>

PLATAFORMA LATTES. **Glossário**. [2022]. Disponível em: <http://lattes.cnpq.br/web/dgp/glossario>. Acesso em: 28 jul. 2022.

ROBINSON, Lyn; PRIEGO, Ernesto; BAWDEN, David. Library and Information Science and Digital Humanities: two disciplines, joint future?. In: PEHAR, Franjo; SCHLOGL, Christian; WOLFF, Christian. Re: inventing Information Science in the Networked Society. **Proceedings of the 14th International Symposium on Information Science (ISI 2015)**, Zadar, Croatia, 19th-21st May 2015. Glückstadt: Verlag Werner Hülsbusch, p. 44-54, 2015. Disponível em: <http://openaccess.city.ac.uk/11889/>. Acesso em: 10 out. 2018.

SANTAREM SEGUNDO, José Eduardo. Web semântica, dados ligados e dados abertos: uma visão dos desafios do Brasil frente às iniciativas. **XVI Encontro Nacional de Pesquisa em Ciência da Informação**, João Pessoa, v. 8, n. 2, p. 219-239, 2015. Disponível em: <http://www.brapci.inf.br/index.php/article/view/0000019443/145a32937c4869ee44422a6ded718148>. Acesso em: 17 set. 2018.

SCHREIBMAN, Susan; SIEMENS, Ray; UNSWORTH, John. **A Companion to Digital Humanities**. Oxford: Blackwell, 2004. Disponível em: <http://www.digitalhumanities.org/companion/>. Acesso em: 26 set. 2018.

SOUSA, Maria Clara Paixão de. Humanidades Digitais: um breve panorama. **Grupo de Pesquisas Universidade de São Paulo**. 2011. Disponível em: <http://www.nehilp.org/~nehilp/HD/SobreHD.html>. Acesso em 10 out. 2018.

SOUSA, Maria Clara Paixão de. O Corpus Tycho Brahe: contribuições para as Humanidades Digitais no Brasil. **Filologia e Linguística Portuguesa**, São Paulo, v. 16, número especial, p. 53-



93, dez 2014. Disponível em: <http://dx.doi.org/10.11606/issn.2176-9419.v16ispep53-93>. Acesso em 21 set. 2018.

SVENSSON, Patrik. Humanities computing as digital humanities. **Digital Humanities Quarterly**, [s.l.], v. 3, n. 3, 2009. Disponível em:<http://www.digitalhumanities.org/dhq/vol/3/3/000065/000065.html>. Acesso em: 11 jan. 2022.

SU, Fangli; ZHANG, Yin; IMMEL, Zachary. Digital humanities research: interdisciplinar collaborations, themes and implications to library and information science. **Journal of Documentation**, Bingley, v. 77, n. 1, p. 143-161, 2020. DOI 10.1108/JD-05-2020-0072

#### **AGRADECIMENTOS:**

Agradecemos a Rafael Jeferson Pezzuto Damaceno pelo auxílio com o processamento dos dados da Plataforma Lattes; a Alan César Belo Angeluci pela preparação do infográfico, e; ao CNPq, pela Bolsa de Produtividade em Pesquisa #311237/2019-3.

#### **CONTRIBUIÇÃO DE AUTORIA**

Os papéis descrevem a contribuição específica de cada colaborador para a produção acadêmica inserir os dados dos autores conforme exemplo, excluindo o que não for aplicável. Iniciais dos primeiros nomes acrescidas com o último Sobrenome, conforme exemplo.

**Concepção e elaboração do manuscrito:** Mugnaini, R.; Guimarães, S. A. S.; Rocha, E. S. S.

**Coleta de dados:** Guimarães, S. A. S.; Rocha, E. S. S.

**ANÁLISE DE DADOS:** Mugnaini, R.; Guimarães, S. A. S.; Rocha, E. S. S.

**Discussão dos resultados:** Mugnaini, R.; Guimarães, S. A. S.; Rocha, E. S. S.

**Revisão e aprovação:** Mugnaini, R.;

Caso necessário veja outros papéis em: <https://casrai.org/credit/>

#### **CONFLITO DE INTERESSES**

Informar conflitos de interesse: financeiros, pessoais, entre possíveis revisores e editores, e/ou possíveis vieses temáticos. Se não houver, mencionar: Não se aplica. Para mais informações: [https://www.abecbrasil.org.br/arquivos/whitepaper\\_CSE.pdf](https://www.abecbrasil.org.br/arquivos/whitepaper_CSE.pdf)

#### **LICENÇA DE USO** – uso exclusivo da revista

Os autores cedem à **Encontros Bibli** os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a [Licença Creative Commons Attribution](#) (CC BY) 4.0 International. Esta licença permite que **terceiros** remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os **autores** têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

#### **PUBLISHER**

Universidade Federal de Santa Catarina. Programa de Pós-graduação em Ciência da Informação. Publicação no [Portal de Periódicos UFSC](#). As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

#### **EDITORES**

Edgar Bisset Alvarez, Ana Clara Cândido, Patrícia Neubert e Genilson Geraldo.

#### **HISTÓRICO**

Recebido em: 16-08-2022 – Aprovado em: 30-01-2023 - Publicado em: 10-02-2023.

