

# Wikidata Lab e Humanidades Digitais: pensamento computacional e prática laboratorial em plataformas colaborativas

**Liráucio Girardi Júnior<sup>I</sup>**

<sup>I</sup>Universidade Municipal de São Caetano do Sul, Faculdade Cásper Líbero, São Paulo, SP, Brasil;

lira.sociologia@gmail.com; <https://orcid.org/0000-0001-5992-0758>

**João Alexandre Peschanski<sup>II</sup>**

<sup>II</sup>Wiki Movimento Brasil, São Paulo, SP, Brasil;

joalpe@wmnobrasil.org; <https://orcid.org/0000-0002-2352-1787>

**Resumo:** O Wikidata Lab apresenta uma série de encontros interdisciplinares, organizados pelo Wiki Movimento Brasil, voltados para o estudo das potencialidades da plataforma Wikidata, que, juntamente com a Wikipédia, a Wikimedia Commons e a Wikiversidade, entre outras, faz parte do ecossistema de plataformas colaborativas da Fundação Wikimedia. Por meio de um estudo de caso normativo, procuramos entender como os encontros do Wikidata Lab representariam um caminho possível para a pesquisa no campo da Comunicação e Informação e de que maneira seriam capazes de articular a chamada virada computacional nas ciências humanas, o trabalho em plataformas colaborativas e os projetos em Humanidades Digitais. Entre as características que identificamos nessas atividades estão o uso do formato “laboratório”, o trabalho interdisciplinar, o desenvolvimento de recursos abertos e colaborativos e a constituição de uma comunidade de práticas que envolvem pesquisadores e pesquisadoras interessados em explorar as consequências dessa virada computacional nas ciências humanas.

**Palavras-chave:** arquitetura de informação; ambiente informacional digital; cultura livre; humanidades digitais

## 1 Introdução

Os campos sociais e, entre eles, o campo acadêmico, são universos relativamente autônomos e objetivados, o que significa que são formados em um processo de depuração mais ou menos prolongada durante sua história. Esse processo não ocorre sem lutas simbólicas, negociações e disputas em torno do desenvolvimento de novos objetos de interesse e novas metodologias, novas regras de organização

do campo e suas respectivas instâncias de legitimação (Bourdieu, 1996, 2001; Martelatto; Pimenta, 2017). Como observa Kirschenbaum (2012), muitas vezes, as “Humanidades Digitais” acabam funcionando como um termo “móvel e tático” para definir uma área capaz de articular diversos agentes como universidades, editoras, associações, órgãos de fomento, imprensa e pesquisadores/pesquisadoras.

Entre as organizações, historicamente, envolvidas com as Humanidades Digitais, encontramos centros de Pesquisa da Universidade de Lisboa, Harvard, Stanford, Maryland, King’s College London, Oxford, Sheffield. Existem, ainda, diversas organizações que têm procurado estimular o desenvolvimento de pesquisas nessa área como ADHO, SDH-SEMI, CenterNet, entre outras.<sup>1</sup>

Entre as experiências latino-americanas, podemos destacar os trabalhos dos laboratórios *Red HD*, da Universidade Autónoma de México, a *Red Colombiana de Humanidades Digitales* (RCHD), o *Laboratório em Rede de Humanidades Digitais* (Larhud) do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), o *LabCult* da Universidade Federal de Minas Gerais (UFMG), o LABHD-UFBA - *Laboratório de Humanidades Digitais* da Universidade Federal da Bahia (UFBA), o *Laboratório de Humanidades Digitais* (LHuD) da Fundação Getúlio Vargas (FGV), o *Laboratório Virtual de Humanidades Digitais* (LaViHD) da Universidade de São Paulo (USP), as parcerias com o *Laboratório de Humanidades Digitais* (Lab-HD), associado à NOVA-FCSH da Universidade NOVA de Lisboa, o *Humanidades Digitales CAICYT Lab*, na Argentina, entre outros.

Neste artigo, não vamos nos concentrar no processo de institucionalização de um Laboratório em Humanidades Digitais propriamente dito, mas em um trabalho laboratorial centrado em encontros de experimentação com os recursos da plataforma Wikidata: o Wikidata Lab<sup>2</sup>.

Embora a plataforma Wikidata seja mais uma entre outras plataformas desenvolvidas pela Fundação Wikimedia (Wikipédia, Wikimedia Commons, Wikiversidade, etc.), o propósito da sua criação está relacionado à organização de um banco de dados estruturados, livre e colaborativo, legível tanto por máquinas quanto por humanos, que permite a criação automatizada de caixas de

informações e a integração entre verbetes da Wikipédia em diversas línguas.<sup>3</sup> Ela encontra-se aberta a desenvolvedores por meio de APIs, o que poderia integrá-la a um modo particular de plataformização (Peschanski, 2021).

A sua criação foi conduzida pelo capítulo alemão da Fundação Wikimedia e financiada pelo *Allen Institute for Artificial Intelligence* (co-fundador da Microsoft), da *Gordon and Betty Moore Foundation* (Gordon Moore foi o cofundador da Intel) e do *Google Inc.* (subsidiária da holding *Alphabet* desde 2015) (Rosen, 2012).

Para a análise dos encontros do Wikidata Lab – e sua possível integração às práticas das Humanidades Digitais – será utilizado o método de estudo de caso normativo.

## 2 O campo das Humanidades Digitais

Embora o termo Humanidades Digitais seja considerado tático e móvel e possa assumir diferentes acepções (na universidade, nas editoras, entre financiadores de pesquisa, na imprensa, etc.), a natureza dos objetos digitais está mudando, consideravelmente, o modo pelo qual fazemos pesquisa (Omena, 2019). Não se trata apenas do *uso* do computador para pesquisas em humanidades, mas da concepção e a utilização de modelagens dos dados em humanidades que permitam um novo entendimento sobre as práticas de representação baseadas na lógica computacional (Unsworth, 2002; Schreibman; Siemens; Unsworth, 2016).

McCarthy (1998) observa que uma ferramenta pode parecer um simples objeto na mão de alguém que não domina o seu significado e o seu uso, mas o seu domínio por um hábil artesão, por exemplo, chega a transformá-la em uma *prótese mental*, um *modo de percepção* e um *instrumento do pensamento* e da ação.

Na maioria das vezes, a área de Humanidades Digitais tem no desenvolvimento de ferramentas e protótipos de software um elemento unificador (e complicador) dos saberes, representando aquilo que Berry (2011) chamou de *virada computacional*, ou seja, um novo modelo epistemológico que reconfigura o conhecimento das ciências humanas a partir da construção de objetos de pesquisa centrado em dados.

Essas novas maneiras de se lidar com os dados em formatos digitais – e com o tipo de modelagem e recursos a que se prestam – têm trazido uma série de questionamentos sobre a formas construção do conhecimento nas mais diversas áreas e têm sido acompanhadas, ao mesmo tempo, por uma profunda crise nos formatos institucionais dos departamentos universitários na sociedade contemporânea.<sup>4</sup>

Esse momento de transformação demanda um conjunto de abordagens específicas capazes de levar em consideração um novo tipo de “ontologia” dos objetos estudados. Ele demanda, como observa Fuller (2008), *uma nova imaginação sociológica* que enfrenta o desafio que a “computacionalidade” dos objetos de estudos traz para a nossa compreensão do que é o conhecimento e, até mesmo, o significado da própria noção de *humanidade*. A possibilidade de lidar com objetos digitais, que estão profundamente imbricados na lógica computacional, faz com que não se esteja falando apenas de novas técnicas, mas de uma profunda reformulação no modo de construção do conhecimento (Berry, 2011).

Nesse sentido, o alinhamento entre computação, ciências da informação e as humanidades não produz apenas novas possibilidades técnicas para a produção de pesquisa, mas reconfigura o próprio modo de construção de problemas. Além disso, acaba rompendo com algumas características da pesquisa nas Humanidades ao enfatizar o trabalho colaborativo, a interdisciplinaridade,<sup>5</sup> a experimentação e a mudança de escala no uso dos dados (McCarthy, 1998; Afanador-Llach *et al.*, 2020).

As Humanidades Digitais podem ser consideradas, por isso, “híbridas, mestiças e impuras” (Fiormonte, 2018), pois reúnem pesquisadores da área das humanidades, ciências da computação, ciências da informação e comunicação e desenvolvem pesquisas na área de: (a) representação visual de dados, (b) representações 3D de projetos arquitetônicos e urbanos de cidades históricas, (c) digitalização de acervos de bibliotecas e museus,<sup>6</sup> (d) desenvolvimento e aperfeiçoamento de linguagens que possibilitem a criação de *templates*, programas, algoritmos, scripts abertos e colaborativos, ajustados aos propósitos das humanidades digitais, (e) formação de pesquisadores em *design* de interfaces

e voltados para a criação de projetos educacionais e artísticos e (f) são multipropósitos e inclusivos, modelos experimentais inspirados nos laboratórios voltados às artes, design e à arquitetura.

Nesse novo ambiente de pesquisa, os humanistas são obrigados a reconfigurar o seu trabalho que passa a ser, também, o de desenvolvedores de “ferramentas” digitais para a produção do conhecimento e experimentação em sua área. Como observa Warwick (2015), o entendimento da condição digital dos objetos aparece, claramente, quando trabalhamos com eles, e isso implica, inclusive, no modo pelo qual será conduzida a formação acadêmica dos novos pesquisadores/pesquisadoras.

O que deve ser incorporado a essa formação não são os novos meios digitais, em si, mas uma cultura a partir da qual esses meios passam a fazer parte da nossa imaginação. A linguagem dessas linhas de pesquisa depende do conhecimento de formatos e linguagens de programação (XML, BIS, R, CSS, C), visualizações de dados, modelagem em 3D, codificação e desenvolvimento de aplicativos livres e abertos e imaginando usos institucionais desses métodos e dados (Ramsay; Rockwell, 2012).

Além disso, o desafio não consiste, apenas, na capacidade dos humanistas de pensar em termos computacionais, mas em trazer as pesquisas no campo computacional para as questões epistemológicas que envolvem o conhecimento nas Humanidades. Deixado como está, o pensamento computacional poderia representar “[...] uma visão extremamente otimista, sem perspectiva histórica, positivista e simplista” do mundo (Van Es; Schäfer; Wieringa, 2021). Embora não seja uma prática “coerente” e se apresente, como vimos, de uma forma tática e fluida, McCarthy (1998) questiona se haverá vontade política, recursos ou imaginação suficiente para realizar esse tipo de aventura intelectual.

Destacamos, então, algumas das propostas levantadas no *Manifesto das Humanidades Digitais* de 2011:

### III. Declaração

5. [...] comunidade de prática solidária, aberta, acolhedora e de livre acesso.
6. [...] comunidade sem fronteiras...multilíngue e multidisciplinar.

7. [...] nossos objetivos são o progresso do conhecimento, o reforço da qualidade da pesquisa em nossas disciplinas, e o enriquecimento do saber e do patrimônio coletivo.

8. Convocamos à integração da cultura digital na definição da cultura geral do século XXI.

#### **IV. Orientações**

9. [...] apelo ao acesso livre aos dados e os metadados... devem ser documentados e interoperáveis, tanto técnica como conceitualmente.

10. Somos a favor da divulgação, da circulação e do livre enriquecimento dos métodos, do código, dos formatos e dos resultados da pesquisa (Dacos, 2011).

### **3 Wikidata Lab: um estudo de caso normativo**

A integração entre o pensamento computacional e a prática laboratorial nas ciências sociais ainda é uma área em desenvolvimento e, nos últimos anos, tem sido enquadrada como práticas típicas do que chamamos de Humanidades Digitais e essa integração passa, necessariamente, pela busca por possíveis novas práticas metodológicas (Dobson, 2019).

Um desafio para pesquisas sobre objetos das Humanidades Digitais é que, por um lado, elas se movimentam entre aquilo que *são*, neste exato momento, e, por outro, pelo que, eventualmente, *virão a ser*, uma vez que se trata de um tipo particular de estudo em ato (Clement, 2016; Van Es; Schäfer; Wieringa, 2021). Há um deslocamento considerável das abordagens que orientam o campo. Além disso, é preciso escapar de abordagens que expõem, acriticamente, características positivas de casos promissores, ou apenas enfatizam as falhas como sua única realidade, destacando suas potencialidades como ilusórias (Wright, 2011).

Nesse sentido, entendemos que o estudo de caso dos encontros do Wikidata Lab tem uma dupla função metodológica. Em primeiro, permite a investigação empírica dos conceitos mais gerais do pensamento computacional e da prática laboratorial. Em segundo, potencializa o próprio processo de construção de categorias de análise, uma vez que se encontram em um campo de possíveis da pesquisa. Trata-se de um esforço dialético direcionado ao estudo de caso normativo (Thacher, 2006) como uma alternativa às compreensões exclusivamente interpretativas ou causais de estudo dos objetos das ciências sociais.

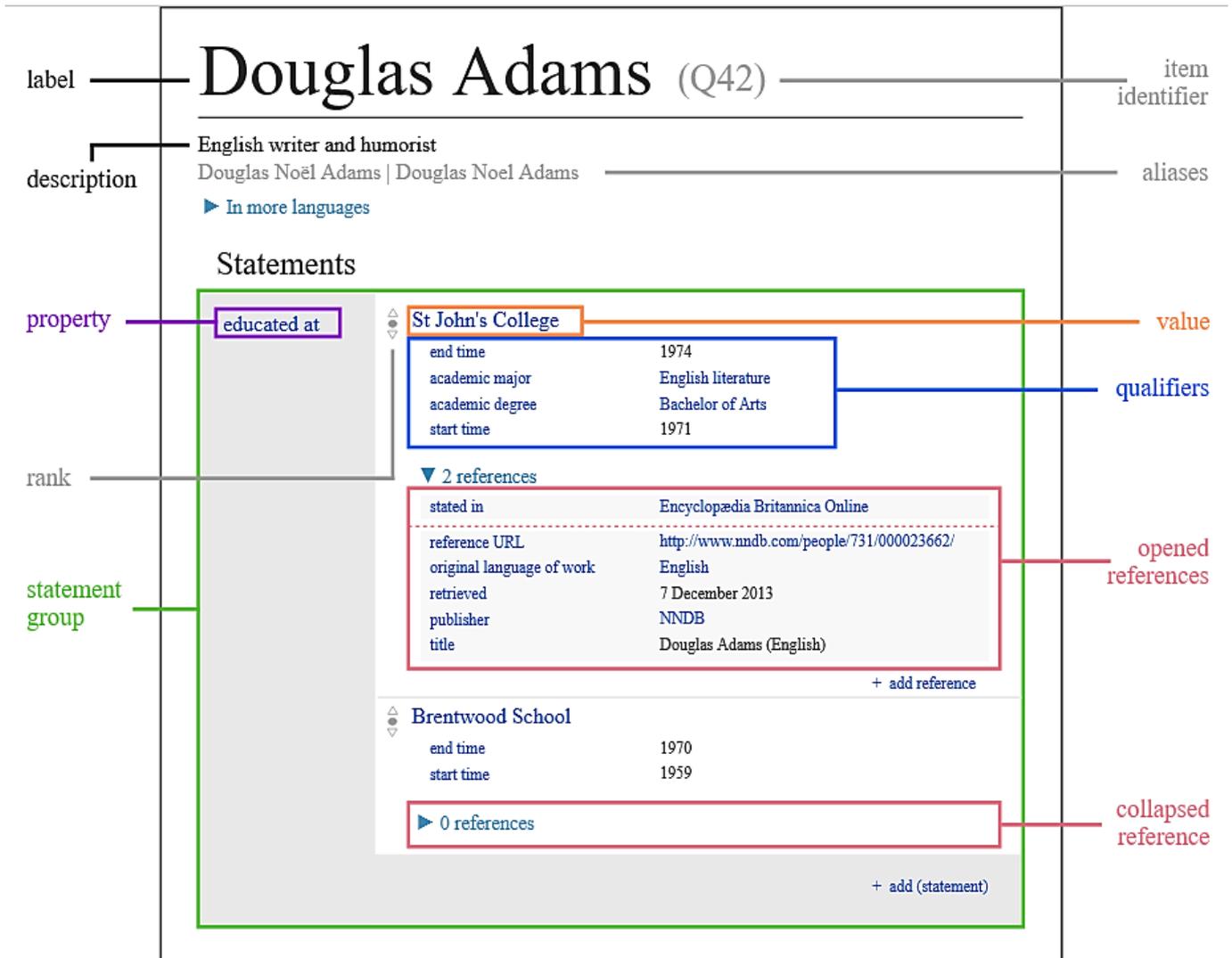
O estudo de caso normativo é uma abordagem metodológica heterodoxa. A pesquisa torna-se ao mesmo tempo descritiva e avaliativa, em contraponto à tendência de separar o elemento abstrato, determinado previamente por formulações teóricas, e a operacionalização no campo. Um corolário é o abandono ao marco teórico apriorístico, buscando, portanto, incorporar as referências à medida que permitem conceituar o objeto. A abordagem remete a linhas recentes nas ciências sociais de desenvolver as formulações teóricas na medida em que surgem necessidades de conceituação (Burawoy, 2009; Charmaz, 2009).

O método adotado neste artigo consistiu no duplo exercício que envolve o exame detalhado de uma prática e o modo pelo qual ela pode ser vinculada aos tipos de abordagens propostas pelas Humanidades Digitais. Será apresentada, de modo sucinto, a sistematização de algumas características descritivas identificadoras do objeto de estudo, seguida do duplo trabalho de definição do Wikidata Lab como um modelo específico de uso do pensamento computacional e prática laboratorial nas ciências humanas e nos estudos sobre plataformas colaborativas.

Lançado em 2012, a plataforma Wikidata é um dos mais novos projetos da Fundação Wikimedia. Trata-se de um ambiente colaborativo criado para o carregamento e edição de dados estruturados e interligados nos verbetes-artigos da Wikipédia, nos carregamentos de imagens, sons e vídeos no Wikimedia Commons etc. Esses dados e os recursos de análise desenvolvidos pelo Wikidata são disponibilizados e mantidos sob licença livre.

Nesse repositório, os elementos são, normalmente, classificados por Identificadores-Q Wikidata, ou QID e descritos, em diversas línguas, por propriedades e valores específicos (Kaffee *et al.*, 2017). Um QID é uma sequência numérica, precedida pela letra Q, única para cada informação armazenada no Wikidata. A maior parte das informações contidas no Wikidata é semântica e estruturada, o que permite a conexão automatizada entre elas, deixando poucos espaços para a produção de categorias de modo livre (Luz; Coneglian; Segundo, 2019).

Figura 1 - Diagrama de um item Wikidata



Fonte: Kritschmar (2016).

O Wikidata Lab promoveu uma série de encontros-eventos-oficinas de compartilhamento de práticas relacionadas ao uso do pensamento computacional e da prática laboratorial centrada para análise e desenvolvimento da base de dados colaborativa do Wikidata. Nesses eventos, reúnem-se acadêmicos, técnicos, estudantes e ativistas do conhecimento livre, dispostos a compartilhar seus projetos e a desenvolver reflexões sobre o significado dos seus trabalhos para o desenvolvimento do pensamento computacional no ecossistema de plataformas Wikimedia.

Os primeiros encontros tiveram início em outubro de 2017, a partir da atuação do Centro de Pesquisa, Inovação e Difusão em Neuromatemática (CEPID NeuroMat) (CEPID NeuroMat, 2013) nas plataformas da Fundação Wikimedia, em parceria com o seu afiliado brasileiro, o *Wiki Movimento Brasil* (Santos, 2019).

O formato desses encontros consiste em apresentação temática inicial, seguida de discussão entre os participantes e atividades práticas. Estas atividades procuram criar ferramentas, projetos e treinamento aos wikipedistas, acadêmicos e wikimedistas envolvidos com projetos GLAM (*Galleries, Libraries, Archives and Museums*), ou seja, projetos relacionados à transferência e compartilhamento de acervos diversos para as plataformas colaborativas da Fundação Wikimedia (Wikipédia, Wikidata, Wikimedia Commons etc.), ampliando a capacidade de acesso, categorização e análise desses recursos.

Entre 2017 e 2020, foram realizados 25 encontros do Wikidata Lab. Cada um dos eventos teve um tema norteador específico, com palestrantes engajados na comunidade Wikimedia, constituindo o que Alves (2016) classificou como uma “comunidade de práticas abertas”. O nível desses encontros não era considerado introdutório às características da plataforma Wikidata, pois envolviam a capacitação de seus membros em pontos avançados dessa tecnologia como, por exemplo, a produção de robôs e o teste de ferramentas ainda em desenvolvimento.

Oficialmente, esses encontros surgiram como um “treinamento para a integração do Wikidata com a Wikipédia em português”, mas o projeto inicial passou por alternância entre sessões voltadas para essa integração e o desenvolvimento de ferramentas específicas de automação no Wikidata. A introdução do pensamento computacional nos processos de edição da Wikipédia não foi recebida com muita tranquilidade pela comunidade de wikipedistas, particularmente, a partir de 2017, pois alguns dos editores da enciclopédia entendiam que essa integração afetava os próprios valores do projeto de edição colaborativa por agentes humanos

A polêmica estava centrada na variedade de ações automatizadas que poderiam ser geradas a partir das ferramentas desenvolvidas com base no

Wikidata. Essas ações poderiam estar relacionadas a funcionalidades pouco controversas como, por exemplo, a conexão de assuntos equivalentes entre vários projetos desenvolvidos nas Wikipédias do mundo todo (Ta; Anutariya, 2014). No entanto, elas também envolviam usos mais controversos – alguns consensualmente proibidos em algumas versões idiomáticas da Wikipédia – tais como a produção automatizada de conteúdo a partir de robôs desenvolvidos na plataforma Wikidata.

O trabalho de estruturação dos dados da Wikipédia no mundo todo foi fundamental para o desenvolvimento de um sistema automatizado de produção de caixas informativas (Sáez; Hogan, 2018), listas automáticas (Warkotsch, 2018) e até textos em linguagem natural (Azzellini; Peschanski; Paixão, 2019). Em princípio, os encontros do Wikidata Lab, desenvolvidos, como vimos acima, pelo Wiki Movimento Brasil, foram pensados para facilitar a integração entre as plataformas Wikidata, Wikimedia Commons e Wikipédia, o que lhes fez receber o prêmio na categoria “Outreach” da conferência WikidataCon, em 2019.

A oitava sessão do Wikidata Lab, por exemplo, foi realizada durante um fim de semana, com uma atividade presencial e contou com a participação de dois renomados contribuidores do Wikidata: Andrew Lih (Fuzheado) e Magnus Manske, respectivamente o responsável pelo setor de Wikimedia no *Smithsonian Institute* nos Estados Unidos e o principal desenvolvedor do *MediaWiki*, o software de base dos projetos Wikimedia, além de algumas das mais importantes ferramentas do Wikidata. Os encontros do Wikidata Lab começaram a ser transmitidos e publicados em vídeo, apenas a partir dessa sessão.

Vamos fazer, agora, um breve levantamento de um conjunto de atividades que caracterizaram os eventos realizados entre 2017 e 2020.

O Wiki Movimento Brasil, uma associação brasileira sem fins lucrativos, fundada em 2013, que é afiliada e tornou-se uma organização oficialmente afiliada da Fundação Wikimedia no Brasil, foi responsável pela idealização do projeto e esteve presente na organização de quase todos esses eventos. Ela esteve ausente em apenas um deles, sendo que as sessões 22 e 26 foram organizadas exclusivamente por esse grupo.

Vinte, dos 26 laboratórios, tiveram a organização do CEPID NeuroMat, estabelecido em 2013, no Instituto de Matemática e Estatística da Universidade de São Paulo e financiado pela FAPESP) (CEPID NeuroMat, 2013), sendo que, em dois deles (sete e nove), o centro foi o único organizador. Todas as atividades que contaram com o CEPID NeuroMat tiveram apoio da FAPESP.

O Museu Paulista da USP participou da organização de quatro eventos e o grupo de pesquisa Tainacan, de um, sendo que as atividades desenvolvidas neles, estava centrada na iniciativa Wiki do Museu do Ipiranga. A plataforma Tainacan foi desenvolvida pelo Laboratório de Inteligência de Redes da Universidade de Brasília (L3P), com apoio da Universidade Federal de Goiás, do Instituto Brasileiro de Informação em Ciência e Tecnologia e do Instituto Brasileiro de Museus, e está voltada para a gestão de coleções/acervos.

O Wikidata Lab contou, também, com apoios de Goethe-Institut (Azzellini, 2020), USP, Fundação de Apoio à Universidade de São Paulo (FUSP) e Fundação Banco do Brasil, além de uma sessão apoiada pelo Wikicite, uma organização que busca melhorar o sistema de referenciamento nos projetos Wikimedia Commons.

A maior parte dos encontros do Wikidata Lab estão voltados para o aperfeiçoamento dos processos computacionais de desenvolvimento do projeto GLAM:

GLAM-Wiki, sigla de *Galleries, Libraries, Archives, and Museums* (galerias, bibliotecas, arquivos e museus) acompanhada do sufixo wiki, que identifica o conjunto de projetos digitais colaborativos e livres sob a guarda da Fundação Wikimedia. [...] O principal trabalho é, indubitavelmente, a enciclopédia eletrônica Wikipédia[...]. Porém, a atuação GLAM-Wiki espalha-se por várias outras plataformas: o repositório de mídias Wikimedia Commons, a base de dados Wikidata e a biblioteca Wikisource, além de espaços de difusão (Outreach Wikimedia) e organização (Meta Wikimedia) (Peschanski, 2021, p. 3).

De 2017 a 2019, as edições foram marcadas por palestras realizadas a partir de uma parceria entre o Wiki Movimento Brasil e o CEPID NeuroMat.

A primeira sessão foi ministrada pelo usuário Ederporto,<sup>7</sup> em 5 de outubro de 2017, que apresentou sobre as “Infocaixas automáticas”, ou seja, quadros que possuem informações estruturadas de forma simples e organizada, que ficam na

lateral superior direita dos verbetes, sintetizando alguns dados de identificação do tópico daquele artigo. Na sessão seguinte, o usuário Mike Peel mostrou em 16 de novembro do mesmo ano como são elaboradas as “Listas do Wikidata usando Listeriabot”, um processo que permite a elaboração semi-automatizada de listas, por exemplo de obras de um acervo.

No mesmo ano, o usuário Sturm apresentou em sete de dezembro de 2017 sobre “Como propor e usar identificadores de controle e autoridade”, que permite a associação entre itens do Wikidata e outras bases de dados notórias, para centralizar fontes diversas em um mesmo espaço.

Em 2018, houve oito Wikidata Labs. Ederporto apresentou em 22 de fevereiro o tópico “Adicionando um grande volume de dados com Quickstatements”, uma interface web que permite aos usuários preparar e importar dados em formatos específicos, como CSV, e vincular itens, tornando o processo de wikidatificação mais rápido.

Ari Hayrinen abordou, em 22 de março, “Uploads em massa com GLAMpipe” (*Graphic data tool for GLAM materials*), uma ferramenta de preparação de dados que buscava simplificar o processo de carregamento em massa, capaz de reconhecer uma série de formatos (texto, imagem, áudio, etc.) e ajustá-los aos padrões do Wikidata.

SandraF (WMF) tratou em 10 de abril do tema: “Dados estruturados no Commons”. O Wikimedia Commons é um repositório central para arquivos de mídia, como imagens, vídeos, áudios e dos mais variados formatos de conteúdo multimídia e a estruturação de dados garante a modelagem de metadados associados aos arquivos de mídia (como informações sobre autor, licença, data, localização, etc.).

Sturm analisou em 17 de maio o tema, “Dados estruturados em mapas interativos”, explicando como informações cartográficas padronizadas podem ser incorporadas e utilizadas na combinação entre a tecnologia livre do Open Street Map (um mapa mundial editável com produção colaborativa de dados geoespaciais abertos) e o Wikidata.

Em nove e dez de junho, Fuzheado e Magnus Manske apresentaram, respectivamente, os temas: “Descobrimo e modelando conhecimento no

Wikidata” e “Imagem, metadados e gamificação do Wikidata”. Tratou-se de uma sessão especial do Wikidata Lab, com o convite a dois palestrantes estrangeiros numa sessão presencial na USP. Além da própria formação da comunidade participante, foi uma etapa de integração mais ampla dos encontros do Wikidata Lab no Movimento Wikimedia, inclusive com a transmissão ao vivo das palestras.

Em nove de agosto de 2018, a usuária EricaAzzellini apresentou sobre “Narrativas estruturadas na Wikipédia”. A proposta foi desenvolver predefinições de linguagem natural para a produção semi-automatizada de artigos na Wikipédia. O usuário Mike Peel fez uma apresentação em 17 de setembro intitulada, “Escrevendo bots no Wikidata”, em que mostrou um passo-a-passo para a produção de scripts para a edição em massa nos projetos da Wikimedia. Em 29 de outubro, Joalpe tratou da “Estruturação de projetos”, utilizando o Wikidata para a organização de iniciativas culturais e educacionais na Wikimedia.

O ano de 2019 contou com oito encontros do Wikidata Lab, começando em 21 de fevereiro com “Reconciliação de metadados com OpenRefine”, apresentado por GiFontelle. A sessão apresentou uma ferramenta livre para ajustes em metadados, especialmente para a doação de conteúdos em massa no Wikimedia Commons. Em 21 de março, Mike Peel falou sobre “Restringindo dados estruturados”, uma prática para manter a consistência, a integridade e a qualidade dos dados.

Na sequência, em 25 de abril, EricaAzzellini explorou “Modelagem com estruturação intencional”, uma sessão na qual analisou a modelagem de metadados para a produção avançada de conteúdos em linguagem natural. Em 23 de maio, Lea Lacroix (WMDE) apresentou sobre “Dados lexicográficos”, a estruturação de informações linguísticas no Wikidata.

Em 13 de junho, Ederporto e GiFontelle apresentaram sobre “Wikiprojetos e métricas de qualidade”, em que mostraram estratégias para controlar os resultados e impactos de projetos culturais no Wikidata, por exemplo a completude e a correção de metadados. Em 11 de julho, a usuária Myxomatosis0 ofereceu o treinamento “Reverso técnicas de scrapping”, para a utilização de *scripts* de coleta de informações estruturadas e consequente contribuição em massa no Wikidata.

Em 8 de agosto, SandraF (WMF) apresentou sobre “Dados estruturados no Wikimedia Commons”, as estratégias para associar informações legíveis por máquina a conteúdos de mídia. Em 10 de setembro, o usuário Denny – o criador do Wikidata – trouxe reflexões sobre “Diversidade de conhecimento e dados conflitantes”, para a interoperabilidade de conteúdos provenientes de fontes heterogêneas.

Houve seis encontros do Wikidata Lab em 2020. Ederporto abriu o ano com “Referências bibliográficas estruturadas”, em 4 de abril, sobre a inserção semi-automática de itens de publicações acadêmicas no Wikidata e sua integração a outros projetos da Wikimedia. Vinte dias depois, o usuário TiagoLubiana apresentou o “Wikiprojeto COVID-19”, sobre os desafios de modelagem sobre a covid-19 no Wikidata e o impacto da transclusão de dados para outros projetos da Wikimedia.

Em 21 de maio, a usuária Esh77 compartilhou conhecimento sobre “Wiki-educação e alfabetização em dados”, uma reflexão sobre a necessidade de letramentos que envolvem uma reflexão sobre a ontologia dos objetos digitais e o pensamento computacional nos âmbitos educacionais, especialmente na universidade.

Em 23 de julho, Ederporto encerrou o ano com a exposição sobre “Posicionamento digital relativo”, sobre a modelagem de imagens no Wikidata com base em protocolos baseados no modelo de Estrutura Internacional para Interoperabilidade de Imagens (IIIF). Em 22 de outubro, o mesmo usuário destacou a “Interoperabilidade e acesso a dados”, em que se discutiu as possibilidades de retroalimentação entre as plataformas Wikimedia e a plataforma de código aberto Tainacan.

Em oito de dezembro, ocorreu uma conversa entre Mike Peel, RexttS, Pigsonthewing, Adamant.pwn e Ederporto sobre o “Módulo de citações estruturadas na Wikipédia”, centrada em um sistema de geração semi-automática de referências na Wikipédia com base nos itens sobre publicações no Wikidata.

Foram desenvolvidos 13 encontros em português e 12, em inglês, nos Wikidata Lab. A predominância da língua inglesa nas últimas edições tem

representado um processo de internacionalização acentuada dos eventos. Houve uma diversidade de nível de instrução e área de formação entre os 31 palestrantes.<sup>8</sup>

Entre 2017 e 2020, nove participantes tinham doutorado e, quatro, mestrado. Dezesete apresentadores eram graduados e, uma, era graduanda. Dada a área de formação, os palestrantes tinham formação em: Matemática (7), Ciência da Computação (6), Comunicação Social (6), Astronomia (4) e História/História da arte (4), além de Biologia (1), Biblioteconomia (1), Educação (1) e Sociologia (1).

O número de visitantes das páginas dos encontros foi regular, com variações esporádicas, possivelmente relacionadas ao interesse pelo tema a ser desenvolvido ou o prestígio do palestrante convidado. Foram consideradas “participantes” as pessoas que se registraram na página de inscrição do evento. A mediana de participantes foi 11 e, a média, 12.

Houve significativa variação nas visualizações dos encontros do Wikidata Lab<sup>9</sup>. As sessões de 23 a 26 tiveram transmissão pelo canal do YouTube e pela página no Facebook do *Wiki Movimento Brasil*, tendo alcançado, no agregado, entre 220 e 360 acessos. As edições oito a 22, excetuando as dez e 13, tiveram transmissão por YouTube apenas, atingindo entre 75 e 305 visualizações, respectivamente. As outras sessões não foram publicadas em vídeo.

Em cada encontro do Wikidata Lab, houve uma atividade prática, direcionada aos participantes, envolvendo a criação de artigos e o de melhorias em artigos ou itens já existentes. Percebe-se que o foco principal dos laboratórios é, primordialmente, a busca pela melhoria de conteúdo: há apenas cinco sessões com a criação de mais de 100 artigos ou itens. A situação mais frequente, com 12 casos, é a de Wikidata Lab com um a 100 criações e entre 101 e 1000 melhorias de conteúdo. No caso em que não houve nem criação nem melhoria é um no qual os participantes desenvolveram juntos um tutorial para uma atividade para usar o Wikidata em sala de aula.

#### 4 Considerações finais

Como destacamos no início do artigo, os campos sociais são resultado de um conjunto de disputas em torno de objetos, metodologias, processos de

institucionalização e de criação de instâncias de legitimação complexas. Em sua origem, as Humanidades Digitais foram identificadas como uma classificação “tática e móvel” por Kirschenbaum (2012), em torno da qual, foram sendo construídos procedimentos de pesquisa, projetos e ambientes (laboratórios) que procuraram escapar à lógica departamental do campo universitário.

Essa proposta demandava novos tipos de considerações sobre o aparecimento dos objetos digitais, uma nova “ontologia” trazida para o campo das Ciências Humanas, provocando um complexo processo de integração entre a lógica das camadas computacionais dos objetos e as ciências preocupadas com sua camada cultural. Um novo tipo de imaginação acadêmica começou a ser estimulada e preparada em uma série de Laboratórios de Pesquisa e “não-conferências” como os THAT CAMP (*The Humanities and Technology Camp*), realizado em 2008, no George Mason Center for History and New Media, e o THAT CAMP, de 2010, realizado em Paris, que gerou o “Manifesto das Humanidades Digitais” (Dacos, 2011).

São várias as iniciativas e os desdobramentos concebidos desde então. Hoje, é possível encontrar reflexões sobre Humanidades Digitais Negras (Gallon, 2016), Humanidades Digitais Feministas, movimentos de decolonização das Humanidades Digitais, entre outros tópicos de discussão. Alguns temas têm focado na organização de Movimentos pelas Humanidades Digitais Públicas (Earhart, 2015) e na configuração das Humanidades Digitais a partir do Sul, ao lembrar que os projetos colonizadores não são apenas focados no domínio político, industrial e tecnológico, mas, também, em termos de soberania cultural e epistemológica (Fiormonte, 2018).

O foco deste artigo concentra-se entre os anos de 2017 (sua criação) e 2020, em que foram realizados 25 encontros do Wikidata Lab centrados em questões como: Infocaixas automáticas; Listas do Wikidata usando Listeria; Como propor e usar identificadores; Adicionando um grande volume de dados com Quickstatements; Uploads em massa com GLAMpipe; Dados estruturados no Commons; Dados estruturados em mapas interativos; Descobrimo e modelando conhecimento no Wikidata; Imagem, metadados e gamificação do Wikidata; Narrativas estruturadas na Wikipedia; Escrevendo bots no Wikidata;

Estruturação do projeto; Reconciliação de metadados com OpenRefine; Restringindo dados estruturados; Modelagem com estruturação intencional, entre outros temas.

A breve descrição dos tópicos desenvolvidos no Wikidata Lab, feita acima, procura mostrar a integração entre certo referencial teórico (móvel, como foi observado) e suas práticas laboratoriais, uma vez que permitem uma reflexão sobre os certos aspectos da experiência proposta pelo Wikidata Lab:

- a) foco na experiência laboratorial vinculada ou não a instituições (universidades, plataformas, etc.);
- b) é um indicador da virada computacional nas ciências humanas, no qual a modelagem de dados e a lógica computacional desempenham um papel importante;
- c) enfatiza a Interdisciplinaridade, a Colaboração e a Experimentação;
- d) destaca a nova ontologia dos Objetos Digitais, o que demanda abordagens epistemológicas novas e o desenvolvimento de novas metodologias e ferramentas de pesquisa;
- e) envolve a criação de uma comunidade de práticas centrada na colaboração aberta e acolhedora (e, muitas vezes, conflitiva, também) entre usuários, pesquisadores e comunidade;
- f) é uma atividade que envolve diversos agentes de diversas nacionalidades e áreas de conhecimento;
- g) está centrada na produção de reflexões sobre o Acesso Livre aos Dados, aos métodos utilizados nos processos e na divulgação aberta dos resultados obtidos;
- h) estimula o entendimento e a incorporação de elementos da Cultura Digital às práticas acadêmicas das ciências humanas.

Embora não possa ser considerado, exatamente, um Laboratório, o Wikidata Lab desenvolve uma série de práticas laboratoriais organizadas como uma “comunidade de práticas”. A ideia de documentar essas atividades e torná-las abertas à participação da comunidade, divulgando processos e ferramentas, marca um dos fundamentos importantes para a sua aproximação com as

Humanidades Digitais.

As reflexões sobre a integração entre camada computacional e camada cultural das plataformas da Fundação Wikimedia indicam o esforço pelo desenvolvimento de uma imaginação sociológica e comunicacional muito particular. Nesse sentido, as Humanidades Digitais não seriam exatamente um campo acadêmico novo, mas uma forma inicial de produzir algumas reflexões sobre os objetos e a cultura digital no campo das ciências humanas.

### Financiamento

O desenvolvimento do artigo está vinculado à pesquisa de João Alexandre Peschanski, que integra os projetos Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) 2013/07699-0 e 2021/06767-8, além da pesquisa envolvendo João Alexandre Peschanski e Liráucio Girardi Júnior, relacionada ao projeto FAPESP 2021/06902-2.

### Referências

AFANADOR-LLACH, M. J *et al.* Humanidades Digitales “a lo colombiche”: cadáver exquisito de la Red Colombiana de Humanidades Digitales. **Revista de Humanidades Digitales**, Madrid, v. 5, p. 217-235, 2020. Disponível em: <https://doi.org/10.5944/rhd.vol.5.2020.27837>. Acesso em: 5 out. 2023.

ALVES, D. As Humanidades Digitais como uma comunidade de práticas dentro do formalismo acadêmico: dos exemplos internacionais ao caso português. **Ler História**, Lisboa, v. 69, p. 91-103, 2016. Disponível em: <https://doi.org/10.4000/lerhistoria.2496>. Acesso em: 20 jan.2022.

AZZELLINI, É. Wikimedistas em residência integrando instituições GLAM à rede de conhecimento dos projetos wikimedia. *In:* VERAS, L. (org.). **Abre-te código: transformação digital e patrimônio cultural: expansão do acesso via dados abertos**. São Paulo: Goethe-Institut, 2020. p. 445-455.

AZZELLINI, É. C.; PESCHANSKI, J. A.; PAIXÃO, F. J. As potencialidades de narrativas estruturadas para o jornalismo computacional: competências jornalísticas na elaboração de textos gerados com bancos de dados. **Texto Livre**, Belo Horizonte, v. 12, n. 1, p. 138-152, 2019. Disponível em: <https://doi.org/10.17851/1983-3652.12.1.138-152>. Acesso em: 5 out. 2023.

BERRY, D. The computational turn: thinking about the digital humanities. **Culture Machine**, United Kingdom, v. 12, p. 1-22, 2011.

BOURDIEU, P. **Meditações Pascalianas**. Rio de Janeiro: Bertrand Brasil, 2001.

BOURDIEU, P. **Razões Práticas**: sobre a teoria da ação. Campinas: Papirus, 1996.

BURAWOY, M. **The extended case method**. Berkeley: University of California Press, 2009.

CENTRO DE PESQUISA, INOVAÇÃO E DIFUSÃO EM NEUROMATEMÁTICA (CEPID NeuroMat). Universidade de São Paulo, São Paulo, 2013. Disponível em: <https://neuromat.numec.prp.usp.br/>. Acesso em: 5 out. 2023.

CHARMAZ, K. **A construção da teoria fundamentada**: guia prático para análise qualitativa. Porto Alegre: Bookman, 2009.

CLEMENT, T. Where is methodology in digital humanities? *In*: SCHREIBMAN, S.; SIEMENS, R.; UNSWORTH, J. (ed.). **A companion to digital humanities**. Oxford: Blackwell. 2016.

DACOS, M. Manifesto das Humanidades Digitais. **Humanidades Digitais - Universidade de São Paulo**, São Paulo, 26 mar. 2011.

DOBSON, J. E. **Critical digital humanities**: the search for a methodology. Champaign: University of Illinois Press, 2019.

EARHART, A. The digital humanities as a laboratory. *In*: GOLDBERG, D. T.; SVENSSON, P. (ed.). **Humanities and the digital**. Boston: MIT Press, 2015. p. 391-400.

FIORMONTE, D. ¿Por qué las humanidades digitales necesitan al sur? *In*: RIO RIANDE, G. *et al.* (org.). **Humanidades digitales**: construcciones locales en contextos globales: Actas del i Congreso Internacional de la Asociación Argentina de Humanidades Digitales. Buenos Aires: Ed. de la Facultad de Filosofía y Letras Universidad de Buenos Aires, 2018.

FULLER, S. **The new sociological imagination**. London: Sage, 2008.

GALLON, K. Making a case for the black digital humanities. *In*: GOLD, M. K.; KLEIN, L. F. (ed.). **Debates in the digital humanities**. Minneapolis: Minnesota University Press, 2016.

GRAHAM, M. The Problem with Wikidata. **The Atlantic**, Washington, 6 Apr. 2012.

LUZ, L. P.; CONEGLIAN, C. S.; SEGUNDO, J. E. S. Tecnologias da web semântica para a recuperação da informação no Wikidata. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 17, p. 1-20, 2019. Disponível em: <https://doi.org/10.20396/rdbci.v17i0.8651791>. Acesso em: 9 fev. 2023.

KAFFEE, L.-A. *et al.* A glimpse into Babel: an analysis of multilinguality in Wikidata. *In: OpenSym'17: INTERNATIONAL SYMPOSIUM ON OPEN COLLABORATION*, 13., 2017, Galway. **Proceedings** [...]. New York: Association for Computing Machinery, 2017. Disponível em: <https://doi.org/10.1145/3125433.3125465>. Acesso em: 9 jan. 2023.

KIRSCHENBAUM, M. Digital Humanities as/is a tactical term. *In: GOLD, M. K. (ed.). Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, 2012.

KRITSCHMAR, C. (WMDE). File: datamodel in Wikidata.svg. **Wikimedia Commons**, [*s.l.*], 21 jun. 2016. Disponível em: [https://commons.wikimedia.org/wiki/File:Datamodel\\_in\\_Wikidata.svg](https://commons.wikimedia.org/wiki/File:Datamodel_in_Wikidata.svg) . Acesso em: 12 mar. 2023.

MARTELATTO, R. M., PIMENTA, R. M. (org.). **Pierre Bourdieu e a produção social da cultura, do conhecimento e da informação**. Rio de Janeiro: Garamond, 2017.

MCCARTY, W. What is humanities computing? Toward a definition of the field. **Centre for Computing in Humanities**, London, 16 Feb.1998.

OMENA, J. J. **Métodos Digitais: teoria, prática, crítica**. Lisboa: ICNOVA, 2019.

PESCHANSKI, J. A. Variedades de processos de difusão digital colaborativa: descrição e análise de iniciativas GLAM-Wiki no Brasil. **Resgate: Revista Interdisciplinar de Cultura**, Campinas, v. 29, p. 1-28, 2021. Disponível em: <https://doi.org/10.20396/resgate.v29i1.8659966>. Acesso em: 5 out. 2023.

RAMSAY, S.; ROCKWELL, G. Developing things: notes toward an epistemology of building in the digital humanities. *In: GOLD, M. K. (org.). Debates in the digital humanities*. Minneapolis: University of Minnesota Press, 2012. p. 75-84.

ROSEN, R. J. Teaching Wikipedia to write itself. **The Atlantic**, Washington, 2 Apr. 2012.

SÁEZ, T.; HOGAN, A. Automatically generating Wikipedia info-boxes from Wikidata. *In: WWW'18: THE WEB CONFERENCE*, 2018, Lyon. **Proceedings** [...]. New York: Association for Computing Machinery, 2018. p. 1823-1830.

Disponível em: <https://doi.org/10.1145/3184558.3191647>. Acesso em: 9 jan. 2023.

SANTOS, J. C. F. **A ciência aberta e suas (re) configurações: políticas, infraestruturas e prática científica**. 2019. Tese (Doutorado) - Instituto de Geociência, Universidade Estadual de Campinas, Campinas, 2019.

SCHREIBMAN, S.; SIEMENS, R.; UNSWORTH, J. (ed.). **A companion to digital humanities**. Oxford: Blackwell. 2016.

TA, T.H.; ANUTARIYA, C. A model for enriching multilingual Wikipedias using infobox and Wikidata property alignment. *In: Supnithi, T. et al. (ed.). Semantic technology: JIST 2014*. Berlin: Springer Cham, 2014. p. 335-350. Disponível em: [https://doi.org/10.1007/978-3-319-15615-6\\_25](https://doi.org/10.1007/978-3-319-15615-6_25). Acesso em: 9 jan. 2023.

THACHER, D. The normative case study. **American journal of sociology**, v. 111, n. 6, p. 1631-1676, 2006.

UNSWORTH, J. What is humanities computing and what is not? **Jahrbuch für Computerphilologie**, Germany, v. 4, p. 71-83, 2002.

VAN ES, K.; SCHÄFER, M. T.; WIERINGA, M. Tool criticism and the computational turn. A “Methodological Moment” in media and communication studies. **M&K Medien & Kommunikationswissenscha**, Germany, v. 69, n. 1, p. 46-64, 2021.

WARKOTSCH, J. **Developing a graphical user interface for generating Wikipedia lists with Wikidata**. 2018. Tese (Doutorado) - Institute of Computer Science, Freie Universit at Berlin, Berlin, 2018.

WARWICK, C. Building theories or theories of building? A tension at the heart of digital humanities. *In: S. Schreibman, S.; Siemens, R.; J. Unsworth, J. (ed.). A new companion to digital humanities*. New Jersey: Wiley, 2015. p. 538-552. Disponível em: <https://doi.org/10.1002/9781118680605.ch37>. Acesso em: 9 jan. 2023.

WRIGHT, E. O. Real utopias. **Contexts**, Washington, v. 10, n. 2, p. 36-42, 2011.

## **Wikidata Lab and Digital Humanities: Computational Thinking and Laboratory Practice in Collaborative Platforms**

**Abstract:** The Wikidata Lab presents a series of interdisciplinary meetings, organized by the Wiki Movimento Brasil, with the aim of studying the potential

of the Wikidata platform. This platform, in conjunction with Wikipedia, Wikimedia Commons, and Wikiversity, among others, is part of the collaborative platform ecosystem of the Wikimedia Foundation. Using a normative case study, our objective is to understand how Wikidata Labs could offer a direction for research in the field of Communication and Information and how they could articulate the so-called computational turn in the humanities, collaborative work on platforms, and projects in Digital Humanities. Among the characteristics we have identified in the projects of Wikidata Labs are the utilization of the “laboratory” format, interdisciplinary collaboration, the development of open and collaborative resources, and the establishment of a community of practices that engage researchers committed to the exploration of this computational turn in the humanities.

**Keywords:** information architecture; digital information environment; open culture; digital humanities

Recebido: 22/05/2023

Aceito: 01/11/2023

#### **Declaração de autoria:**

**Concepção e elaboração do estudo:** Liráucio Girardi Júnior, João Alexandre Peschanski.

**Coleta de dados:** Liráucio Girardi Júnior, João Alexandre Peschanski.

**Análise e interpretação de dados:** Liráucio Girardi Júnior, João Alexandre Peschanski.

**Redação:** Liráucio Girardi Júnior, João Alexandre Peschanski.

**Revisão crítica do manuscrito:** Liráucio Girardi Júnior, João Alexandre Peschanski.

#### **Como citar**

GIRARDI JÚNIOR, Liráucio; PESCHANSKI, João Alexandre. Wikidata Lab e Humanidades Digitais: pensamento computacional e prática laboratorial em plataformas colaborativas. **Em Questão**, Porto Alegre, v. 30, e-132666, 2024. DOI: <https://doi.org/10.1590/1808-5245.30.132666>



<sup>1</sup> ADHO - *Alliance of Digital Humanities Organizations*; CHAIN - *Coalition of Humanities and Arts Infrastructures and Networks*; CHCI - *Consortium of Humanities Centers and Institutes*; CLARIN - *Common Language Resources and Technology Infrastructure*; DARIAH - *Digital Research Infrastructure for the Arts and Humanities*.

<sup>2</sup> Todos os dados que baseiam as análises e conclusões deste artigo estão tabelados e disponíveis em <https://zenodo.org/doi/10.5281/zenodo.8433117>.

<sup>3</sup> Graham (2012) observa que a plataforma, criada em 2012, procura avançar em direção à Web Semântica, mas que isso traz, também, diversos problemas, como uma tentativa de estruturação de verbetes da Wikipédia que são marcados

---

visões conflitantes de acordo com o país, como modos diversos de estruturação de tópicos que utiliza e com a comunidade linguística em que o conteúdo é desenvolvido para diferentes tipos de leitores. A preocupação em fazer com que a informação seja cada vez mais adequada à compreensão por inteligência artificial deve vir acompanhada, também, de uma análise adequada da diversidade de visões e do caráter agonístico que certos temas podem ter. Este não é o foco do nosso artigo, mas merece uma importante reflexão.

<sup>4</sup> Uma análise crítica sobre a criação do Wikidata pode ser vista em Graham (2012).

<sup>5</sup> A tentativa de manter essas transformações dentro dos departamentos acadêmicos tradicionais tem levado a uma multiplicação de classificações concorrentes, tais como: Digital History, Digital Literary Studies Digital Sociology etc.

<sup>6</sup> Como descrevem Afanador-Llach *et. al.* (2020 p. 219): “En Colombia es precisamente desde instituciones de la memoria y de la gestión del patrimonio cultural, como bibliotecas, archivos y museos, donde varios proyectos de digitalización y apropiación de la cultura digital comenzaron a proponer espacios y oportunidades para pensar y hacer con lo digital.”

<sup>7</sup> O nome dos palestrantes corresponde, algumas vezes, ao nome de usuário utilizado para acesso às plataformas Wikimedia.

<sup>8</sup> Alguns convidados deram mais de uma palestra. Houve, também, sessões com mais de um palestrante.

<sup>9</sup> São considerados apenas os dados analíticos do canal do YouTube e da página no Facebook do Wiki Movimento Brasil até 31/12/2020. Nota-se que em alguns casos vídeos foram transmitidos em outras tecnologias do Wiki Movimento Brasil; também foram publicados em plataformas de outras entidades.