



A transcrição e a leitura de manuscritos entre o crowdsourcing e a participação cidadã

Ana Margarida Dias da Silva^a, Leonor Calvão Borges^b

^a*Faculdade de Letras da Universidade de Coimbra, Portugal,
margaridadiasdasilva@gmail.com*

^b*Faculdade de Letras da Universidade de Coimbra, Portugal, leonorcborges@gmail.com*

Resumo

O principal objetivo deste trabalho é identificar e analisar projetos participativos (e massivos) de transcrição e leitura de manuscritos realizados por instituições de memória, como bibliotecas, arquivos e museus. Pretende-se identificar a(s) tipologia(s) de documentos mais utilizada(s) em projetos colaborativos, analisar os requisitos necessários para a participação e o grau de envolvimento dos utilizadores, cidadãos e comunidades de prossumidores. São analisados as boas práticas e os benefícios na disseminação do conhecimento através da web 2.0 e o contributo do aproveitamento da inteligência coletiva. Conclui-se que existe uma utilização crescente de plataformas web 2.0 por instituições de memória para projetos de crowdsourcing e participação cidadã associados à leitura e transcrição de manuscritos, onde as comunidades de genealogistas e, em menor escala, historiadores da área da história da família desempenham papel relevante. A participação e colaboração de todos permitem um maior grau de exaustividade da descrição arquivística, tornando a participação coletiva fundamental na recuperação da informação, e dando assim um novo uso à paleografia, cujo conhecimento é essencial para a leitura de manuscritos..

Palavras-chave: Paleografia, Participação colaborativa, Web 2.0, Arquivos públicos.

Introdução

A paleografia, ciência que, na sua vertente prática, se dedica ao ensino da leitura de manuscritos antigos, independentemente da sua língua, esteve tradicionalmente associada aos cursos de História e Ciências Documentais, aprendida nas academias, e vista como ciência auxiliar (Marques 1981). A necessidade de ler documentos antigos surgiu desde cedo nas administrações, levando à especialização de profissionais da leitura. Essa realidade vai fazer com que a paleografia esteja inevitavelmente associada à ciência diplomática, surgindo como uma das formas de validação da veracidade dos documentos (Marques 1918; Borges & Silva 2018).

As Tecnologias da Informação e Comunicação (TIC), a *World Wide Web* (WWW) e a Internet trouxeram novas formas de comunicação e de acesso à informação, revolucionaram a forma de relacionamento dos utilizadores, clientes e cidadãos com a informação e o desenvolvimento da web 2.0 transformou-nos a todos, simultaneamente, em prossumidores (consumidores e produtores) de informação (Silva 2013).

No século XXI, as potencialidades da web 2.0 e a participação colaborativa *on-line* são uma realidade que abrange as designadas instituições de memória (arquivos, bibliotecas e museus) e são cada vez mais os arquivos que aderem a plataformas da web 2.0 para potenciar a identificação e comunicação dos seus documentos (Silva 2013; Silva & Borges 2017).

Isto porque “As particularidades do ambiente internet, onde se destacam aspectos como a interatividade, a democratização do acesso às informações, a quebra de barreiras geográficas e

o desenvolvimento da tecnologia de telecomunicações, modificaram não apenas a forma de lidar com padrões de representação e organização da informação até então estabelecidos mas, também, o comportamento humano e a forma de comunicação da sociedade contemporânea.” (Barros 2011: 16).

Crowdsourcing é um neologismo de origem britânica que resulta da agregação dos termos *crowd* (multidão) e *outsourcing* (terceirização) e que pode ser traduzido como “a sabedoria das multidões”. Este conceito foi utilizado, pela primeira vez, em 2006 num artigo de Jeff Howe para a *Wired Magazine* para designar um novo modelo colaborativo de negócios que rentabiliza a criatividade coletiva na *World Wide Web* (Brabham 2009; Estelles-Arolas & González Ladrón de Guevara 2012)

Esta participação pública pode ser vista como uma extensão lógica do processo democrático de formas mais locais, diretas e deliberativas, a que se acrescenta a valorização do conhecimento de não-especialistas pois “Participation is the act of creating new knowledge, contributing new perspectives to the planning process, and diffusing knowledge to others in the process” (Brabham 2009: 244).

A progressiva disponibilização de conteúdos em acesso aberto, de certa forma em consequência da digitalização massiva das coleções existentes em arquivos, bibliotecas e museus, associada à web 2.0 e às diferentes formas de aprendizagem à distância, resultaram no aparecimento de projetos e sites de divulgação e promoção da leitura de manuscritos, que passam, por exemplo, pela indexação social de documentos de arquivo ou pela sua transcrição colaborativa (Silva 2017; Borges & Silva 2018). Nesse sentido, é objetivo deste trabalho é identificar e analisar projetos participativos (e massivos) de transcrição e leitura de manuscritos realizados por instituições de memória, como bibliotecas, arquivos e museus.

Método

Parte-se da revisão da literatura científica sobre a transcrição massiva de documentos e sobre o estabelecimento de novas conexões com os cidadãos através do uso de plataformas web 2.0.

A identificação dos projetos colaborativos foi feita *on-line* em instituições de memória que solicitam a participação na transcrição de manuscritos, não considerando projetos de transcrição de material impresso digitalizado através de OCR, como o *Projeto Gutenberg*¹, mas englobando projetos onde, em formulários pré-estabelecidos (e impressos), os dados resultantes da leitura dos manuscritos são registados, como o *Operation War Diary*².

A pesquisa foi limitada aos Estados Unidos da América, França e Reino Unido, e foi realizada uma análise qualitativa dos projetos identificados. A seleção foi feita com base no elevado número de exemplos e abrangência de projetos, por um lado, e por estarem bem documentados, por outro. Para a análise e discussão dos resultados foram recolhidos os seguintes dados: instituições promotoras, designação e data de início dos projetos, tipologia e âmbito cronológico dos documentos a serem transcritos, método de participação, solicitação (ou não) dos participantes com conhecimento paleográfico, número de participantes, número de documentos transcritos, e bibliografia e informação sobre eles.

Resultados e Discussão

A partir da análise dos projetos identificados infere-se o seguinte:

- Os projetos são desenvolvidos principalmente no início do século XXI, em arquivos nacionais, departamentais, locais ou até em galerias nacionais, no caso francês, mas nos Estados Unidos e

¹ <https://www.gutenberg.org/>

² <https://www.operationwardiary.org/>

no Reino Unido há uma conexão próxima com o ambiente acadêmico, que colabora (ou coordena) projetos neste campo como, por exemplo, o projeto *Transcribe Bentham*³ ou *Caro Professor Einstein: O Comitê de Emergência de Cientistas Atômicos na América do Pós-Guerra*⁴.

- No caso dos arquivos nacionais ingleses e norte-americanos, existe ainda um intencional apelo à transcrição de documentos como forma de melhorar os recursos informativos, uma vez que, como refere o NARA “Together we can make the records of the National Archives more discoverable online”, tornando o utilizador colaborador. Para tal, fazem-se apelos como os do *UK National Archives*, *Become a Citizen Historian*, são criadas páginas destinadas ao “arquivista cidadão”, e fornecidos materiais de aprendizagem de paleografia. Em França, os *Archives Nationales* destacam a importância da leitura colaborativa de documentos em linha pois consideram que não é suficiente o acesso aos documentos digitalizados, mas imperativa e necessária a compreensão do seu conteúdo através do decifrar das escritas paleográficas. É por esse motivo que em França são dinamizados dezenas de cursos de paleografia⁵.

- Nos Estados Unidos esta tendência faz-se sentir noutras instituições de memória, destacando-se o *Smithsonian Museum*, com a criação, em 2013, do *Smithsonian Transcription Center*⁶ especialmente dedicado aos voluntários digitais daquela instituição, com o objetivo de “improve and increase the quality of public engagement with Smithsonian materials, increase access and use of our digitized content, and create pathways of learning and new knowledge between the public and Smithsonian staff”;

- Os projetos são tradicionalmente desenvolvidos em instituições de memória, embora alguns tenham ligação direta com projetos de Humanidades Digitais e, de forma embrionária, com a ciência cidadã, como é visível no uso da plataforma *Zooniverse*⁷. Nestes casos, verifica-se uma maior diversidade de temas e coleções de documentos a transcrever, encontrando-se projetos literários (*Shakespeare’s World*⁸), artísticos (*AnnoTate*⁹), científicos (*Reading Nature’s Library*¹⁰) ou mesmo relacionados com grupos minoritários (*African American Civil War Soldiers*¹¹);

- Alguns dos projetos são realizados e relacionados a importantes eventos históricos, como a Primeira Guerra Mundial. Em França, arquivos nacionais, departamentais e locais, trabalham

³ <http://blogs.ucl.ac.uk/transcribe-bentham/>. Trata-se de um projeto desenvolvido pela *University College London* (UCL), com o objetivo de transcrever colaborativamente a obra completa de Jeremy Bentham.

⁴ <http://scarc.library.oregonstate.edu/omeka/exhibits/show/ecas/about>. Projeto desenvolvido pela *Oregon State University Libraries and Press* para disponibilizar as cartas de cidadãos e cientistas enviadas no *Emergency Committee of Atomic Scientists* na sequência do lançamento da bomba atômica.

⁵ <http://www.culture.fr/Genealogie/Articles/Cours-de-paleographie-en-ligne> Ver também Borges & Silva 2018.

⁶ <https://transcription.si.edu/about>

⁷ <https://www.zooniverse.org/>

⁸ <https://www.shakespearesworld.org/#/>, Projeto de transcrição colaborativa desenvolvido pela *Folger Shakespeare Library* em Washington, a *Universidade de Oxford* e o *Oxford English Dictionary of Oxford University Press*, com o objetivo de tornar acessíveis manuscritos escritos entre 1564-1616, e assim compreender melhor o mundo em que movia o autor. Um dos outputs deste projeto é ainda melhorar o dicionário de língua inglesa, dando assim um verdadeiro contributo para o desenvolvimento do estudo da língua inglesa.

⁹ <https://anno.tate.org.uk/#/>. Projeto de transcrição colaborativa desenvolvido pela *Tate Gallery*, com o objetivo de tornar acessível documentação relativa a artistas britânicos, nomeadamente diários e correspondência.

¹⁰ <https://www.zooniverse.org/projects/mzfasdg2/reading-natures-library/about/research> Trata-se de um projeto diferente, levado a cabo pelo *Manchester Museum* com o objetivo de transcrever as etiquetas da sua importante coleção de geologia, que contém 100,000 fósseis e 40.000 pedras e minerais. Referindo a falta de pessoal da instituição para esta tarefa, apela-se à disponibilidade dos cidadãos para transferir a informação sobre as peças para uma base de dados *on-line*.

¹¹ <https://www.zooniverse.org/projects/usct/african-american-civil-war-soldiers>. Projeto colaborativo que junta historiadores, cientistas sociais e o *African American Civil War Museum*, que o coordena, em conjunto com o *National Archives and Records Administration* (Washington D.C.), *Fold 3* (Salt Lake City, UT) e o *Willson Center Digital Humanities Lab at the University of Georgia* (Atlanta, GA)

juntos e iniciaram a transcrição *on-line* em massa de registos relacionados com a 1ª Grande Guerra com o projeto *Mémoire des hommes*¹². Ainda em França, apela-se à participação de voluntários na *Opération Transcription Saison 2*, projeto colaborativo de transcrição *on-line*, iniciado pela *Ecole Nationale des Chartes* e os *Archives Nationales*, igualmente no âmbito da comemoração do aniversário da Grande Guerra¹³. Também no Reino Unido é dado destaque à identificação e análise da participação daquele país na 1ª Guerra Mundial, com o projeto *Operation War Diary: reports from the front*¹⁴;

- Em relação ao público-alvo, o grupo que responde massivamente às solicitações de transcrição e criação de pontos de acesso são comunidades de genealogistas e, em menor escala, historiadores da área da história da família. Não é de surpreender pois a maioria dos projetos refere-se a registos paroquiais ou outra documentação a partir da qual a identificação pessoal e familiar possa ser derivada, tais como censos populacionais, por exemplo. Em França, a *Société des Amis des Archives de France*, a *Révue Française de Généalogie*, *Les Amis des Archives de Indre*, a *Société Généalogique du Bas-Berry*, a *Association des Amis des Archives de la Haute-Garonne* e a *Société généalogique du Lyonnais et du Beaujolais*, só para citar alguns exemplos, têm papel ativo junto dos arquivos nacionais, departamentais e locais, na promoção da leitura de manuscritos de forma colaborativa. Em Inglaterra, país com forte implantação de comunidades locais, existem inúmeros projetos lançados pelas mesmas ligadas a arquivos, como por exemplo o caso da *Oxfordshire Family History Society*¹⁵. Contudo, é o grande projeto do *Family Search Indexing*¹⁶ lançado pela Igreja Mormom, sediada nos Estados Unidos, mas com projetos nacionais e locais na maior parte dos países europeus, o grande impulsionador e agregador de mais contributos em termos internacionais;

- A documentação a transcrever é maioritariamente textual. Existem, ainda assim, projetos com outro tipo de documentos, como o *Transcribe Sctolands Spaces*¹⁷, da iniciativa da *Royal Commission on the Ancient and Historical Monuments of Scotland*, a *National Library of Scotland* e o *National Records of Scotland*, com o objetivo de transcrever e identificar mapas, fotografias, registos arqueológicos, desenhos e documentos relativos a impostos relativos àquele país. Também o *Old Weather Project*¹⁸, projeto desenvolvido pelos *U.S. National Archives* (NARA), *UK National Archives* e *UK National Meteorological Services*, privilegia registos científicos de meteorologia, entre outros.

- Quanto ao âmbito cronológico da documentação é mais abrangente no caso dos registos paroquiais e, nos demais projetos, circunscritos aos séculos XVIII, XIX e XX;

- A maioria dos projetos exige o registo dos voluntários nas plataformas e, em alguns casos, disponibilizam-se manuais de leitura paleográfica ou algumas noções sobre a história da escrita;

- Dada a falta de dados sobre a maioria dos locais do projeto não é possível estabelecer uma escala percentual para a totalidade do número de voluntários a eles agregados. No entanto, analisando os resultados apresentados noutros tipos de plataformas, como o *Flickr Commons*, verifica-se a existência de *super taggers*, ou, neste caso, *super paleógrafos*.

Em termos gerais, verifica-se que a comunidade científica na área da Ciência da Informação está atenta, reflete e desenvolve novas formas de comunicação com o público, tornando-o ao mesmo tempo consumidor e produtor de conteúdos, nomeadamente, mas não restringido, através da etiquetagem colaborativa (Barros 2011; Bouvé 2012; Moirez 2012; Palmer 2009; Ridge 2017; Silva 2017; Yedid 2013; Zastrow 2014).

¹² <http://www.memoiredeshommes.sga.defense.gouv.fr/fr/article.php?larub=52&titre=indexation-collaborative>

¹³ <http://archives.yvelines.fr/article.php?laref=1574&titre=operation-transcription-saison-2-avis-aux-volontaires->

¹⁴ <https://www.operationwardiary.org/>

¹⁵ <http://searches.oxfordshirefhs.org.uk/parhelp.html>

¹⁶ <https://www.familysearch.org/indexing/>

¹⁷ <https://scotlandspplaces.gov.uk/scotlandsp-places-update>

¹⁸ <https://www.oldweather.org/>

Contudo, e especificamente no que às transcrições massivas de documentos diz respeito, registam-se poucos artigos de âmbito geral (Alcalá Ponce de León), ou, subsidiariamente, de identificação de formação paleográfica *on-line* (Santos Ruiz 2017; Stokes 2015; Zozaya-Montes 2014; Borges & Silva 2018).

Apesar disso, os projetos lançados são objeto de criterioso estudo e divulgação (Causer & Wallace 2012; Grayson 2016; Parilla & Ferriter 2016; Saylor & Wolfe 2011), permitindo a análise dos procedimentos adotados e eventuais adaptações ou melhorias.

As designadas instituições de memória (arquivos, bibliotecas e museus) fornecem, para além das funções de salvaguarda, organização e descrição, acesso à informação. Esta última será, por ventura, a sua função mais importante visto que permite a recuperação de informação relevante quer para a própria organização quer para os investigadores externos (Silva 2017).

Com as plataformas da web 2.0 e os projetos de *crowdsourcing* verifica-se que “Los entornos colaborativos, que facilitan el aprovechamiento de la inteligencia colectiva y de los patrones de uso, han modificado la forma en que los usuarios se relacionan con la información, y también la forma en que los proveedores de servicios de información se relacionan con los usuarios.” (Yedid, 2013: 14). Esta mudança de relação entre os serviços e os utilizadores tem a sua génese na adoção de ferramentas da web 2.0. O *crowdsourcing*, com o convite à participação e colaboração de todos no realizar de funções tradicionalmente realizadas exclusivamente por profissionais da informação, nomeadamente na descrição de conteúdos e criação de pontos de acesso (Silva 2017), é tanto mais relevante quanto “sabemos que o instrumento de pesquisa de maior utilização nos arquivos é o Inventário e que, na prática, é inviável chegar à produção de outros instrumentos mais detalhados para todos os fundos, pelo simples facto de a sua elaboração ser extraordinariamente morosa.” (Ribeiro, 1996: 22).

Assim, no que diz respeito à leitura colaborativa de manuscritos, “La nouveauté réside dans l’ampleur des acteurs concernés et dans l’ampleur du travail qui est mis en commun.” (Bouye, 2012: 1). E a novidade reside também numa mudança de paradigma e de mentalidade pois, “Délaissant souvent les interactions et échanges superficiels avec leurs usagers, ils [os arquivos] développent en revanche des projets ambitieux de crowdsourcing, d’« archives participatives », fondés sur l’apport de connaissances et de compétences des internautes et destinés à améliorer et faciliter l’accès à leurs fonds.” (Moirez, 2012: 188).

Conclusões

Conclui-se que existe uma utilização crescente de plataformas web 2.0 por instituições de memória para projetos de *crowdsourcing* e participação cidadã. Estes projetos de transcrição massiva de documentos permitem que qualquer pessoa com acesso à Internet participe na transcrição e leitura de manuscritos, o que de outra forma seria impensável, dando assim um novo uso à paleografia, cujo conhecimento é essencial para a leitura de manuscritos. A participação e colaboração de todos permitem um maior grau de exaustividade da descrição arquivística, que de outra forma não seria possível nos arquivos, tornando a participação fundamental na recuperação da informação. A comunidade dos genealogistas desempenha papel importante na colaboração destes projetos a nível dos países analisados.

Referências:

Alcalá Ponce de León, M. (2015). Crowdsourcing a les institucions de la memòria: les transcripcions massives. *BiD: Textos Universitaris de Biblioteconomia i Documentació*, (35 (desembre)). Acessível em: <http://bid.ub.edu/es/35/alcala.htm>

Barros, L. M. de S. (2011). A Folksonomia como prática de classificação colaborativa para a recuperação da informação. Universidade Federal do Rio de Janeiro. Acessível em: <http://ridi.ibict.br/handle/123456789/737>

Borges, L. C.; Silva, A. M. D. (2018). Transcrições em linha: *e-learning* de Paleografia em arquivos europeus. *Revista Portuguesa de História*. T. XLVIII, 39-59 – ISSN: 0870.4147 DOI: https://doi.org/10.14195/0870-4147_49_2

Bouyé, É. (2012). Le web collaboratif dans les services d'archives publics: un pari sur l'intelligence et la motivation des publics. In *Rencontre annuelle des services d'archives départementales (RASAD)*, 1–12. Bordeaux. Acessível em: http://archives.cantal.fr/download.cgi?filename=accounts/mnesys_ad15/dat...

Brabham, D. C. (2009). Crowdsourcing the Public Participation Process for Planning Projects. *Planning Theory*. 8: 242-262. Acessível em: DOI: 10.1177/1473095209104824

Causser, T.; Wallace, V. (2012). Building A Volunteer Community: Results and Findings from Transcribe Bentham. *DHQ: Digital Humanities Quarterly*, 6(2). Acessível em: <http://www.digitalhumanities.org/dhq/vol/6/2/000125/000125.html>

Estellés-Arolas, E.; González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*. Vol. 38, nº. 2, 189-200. Acessível em: <http://www.crowdsourcing-blog.org/wp-content/uploads/2012/02/Towards-an-integrated-crowdsourcing-definition-Estell%C3%A9s-Gonz%C3%A1lez.pdf>

Grayson, R. S. (2016). A Life in the Trenches? The Use of Operation War Diary and Crowdsourcing Methods to Provide an Understanding of the British Army's Day-to-Day Life on the Western Front. *British Journal for Military History*, 2(2), 160–185. Acessível em: <https://research.gold.ac.uk/16713/1/Grayson%252c%20R.pdf>

Marques, A. H. de O. (1981). *Paleografia*. In *Dicionário de História de Portugal dirigido por Joel Serrão*. Vol. 4. Porto: Livraria Figueirinhas, 528-534.

Moirez, P. (2012). Archives participatives. *Bibliothèques 2.0 à l'heure des médias sociaux*. [Em linha]. *Editions du Cercle de la librairie*, 187-197. Acessível em: http://archivesic.ccsd.cnrs.fr/docs/00/72/54/20/PDF/ArchivesParticipatives_PMoirez.pdf

Palmer, Joy. 2009. Archives 2.0: If We Build It, Will They Come?. *Ariadne* 60. Acessível em: <http://www.ariadne.ac.uk/issue60/palmer>

Parilla, L., & Ferriter, M. (2016). Social Media and Crowdsourced Transcription of Historical Materials at the Smithsonian Institution: Methods for Strengthening Community Engagement and Its Tie to Transcription Output. *The American Archivist*. 79 (2), 212–234. Acessível em: <http://americanarchivist.org/doi/abs/10.17723/0360-9081-79.2.438?code=same-site>

Ribeiro, F. (1996). *Indexação e Controlo de Autoridade em Arquivos*. Porto: Câmara Municipal do Porto, Ed. Acessível em: <http://hdl.handle.net/10216/10721>

Ridge, M. (2017). The contributions of family and local historians to British History online. In Roued-Cunliffe, H., & Copeland, A. (Eds.). *Participatory heritages*. London: Facet Publishing, 57-66.

Santos Ruiz, V. (2017). *Paleografía digital: reto y necesidad de los profesionales de archivo*. Trabajo de fin de Master en Archivos, Gestión Documental y Continuidad Digital à Universidad Carlos III de Madrid. Acessível em: <http://hdl.handle.net/10016/26013>

Saylor, N., & Wolfe, J. (2011). Experimenting with Strategies for Crowdsourcing Manuscript Transcription. *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC*, (277), 9–14. Acessível em: <http://publications.arl.org/rli277/>

Silva, A. M. D. da; Borges, L. C. (2017). Texto versus Imagens? Folksonomias e Indexação Social em Arquivos. In *Tendências Atuais e Perspetivas Futuras em Organização do Conhecimento: atas do III Congresso ISKO Espanha e Portugal - XIII Congresso ISKO Espanha*. Universidade de Coimbra 23-24 novembro, 391-400. Acessível em: <https://purl.org/sci/atas/isko2017>

Silva, A. M. D. da. (2017). Folksonomies in archives: controlled collaboration for specific documents” [Em linha] *Ariadne* 77 (Junho). Acessível em: <http://www.ariadne.ac.uk/issue77/margaridadiasdasilva>

Silva, A. M. D. da. (2013). O Uso da Internet e da Web 2.0 na difusão e acesso à informação arquivística: o caso dos arquivos municipais portugueses. Universidade Nova de Lisboa. Acessível em: <http://run.unl.pt/handle/10362/12014>

Stokes, P. (2015). Digital approaches to paleography and book history: some challenges, present and future. *Frontiers in Digital Humanities* 2:5. Acessível em: <https://doi.org/10.3389/fdigh.2015.00005>

Yedid, N. (2013). Introducción a las Folksonomías: Definición, Características y Diferencias con los Modelos Tradicionales de Indización. *Información, Cultura Y Sociedad*, 29, 13– 26. Acessível em: <http://www.scielo.org.ar/pdf/ics/n29/n29a02.pdf>

Zastrow, J. (2014). Crowdsourcing Cultural Heritage: “Citizen Archivists” for the Future. *Computers in Libraries*, 34 (8). Acessível em: <http://www.infotoday.com/cilmag/oct14/Zastrow--Crowdsourcing-Cultural-Heritage.shtml>

Zozaya-Montes, L. (2014). Cursos online de paleografía: herencias, limitaciones, logros y propuestas. *El profesional de la información*, septiembre-octubre, v. 23, n. 5, 475-484. Acessível em: <http://dx.doi.org/10.3145/epi.2014.sep.04>