

Transcrições automáticas nos arquivos distritais portugueses: acelerar o acesso à informação*

Ana Margarida Dias da Silva**

Cultura. Revista de História e Teoria das Ideias 41-42 (2023), 269-286. ISSN 0870-4546

URL: <https://revistas.fcsh.unl.pt/cultura/article/view/1226>

Resumo

Desde 1979, os arquivos distritais portugueses iniciaram uma política de digitalização. Contudo, a mera disponibilização de objetos digitais, sem informação associada ao nível do registo, não permite uma pesquisa rápida e eficaz. A Paleografia Digital é um campo fértil para auxiliar arquivos e arquivistas na captação de informação e como ferramenta para facilitar o acesso à informação, nomeadamente através do Reconhecimento de Texto Manuscrito (HTR), mas tem pouca ou nenhuma utilização nos arquivos portugueses. Procura mostrar-se como modelo *Portuguese handwriting 16th-19th c.* pode ser utilizado pelos arquivos distritais portugueses para acelerar o acesso à informação, nomeadamente, nos registos paroquiais.

Palavras-chave: arquivos distritais, Paleografia Digital, Humanidades Digitais, registos paroquiais, memória coletiva.

Abstract

Since 1979, Portuguese district archives have begun a policy of digitisation. However, the mere availability of digital objects, without associated information at the record level, does not allow for quick and effective searching. Digital Palaeography is a fertile field for helping archives and archivists capture information and as a tool for facilitating access to information, particularly through Handwritten Text Recognition (HTR), but it has little or no use in Portuguese archives. The aim is to show how the *Portuguese handwriting 16th-19th c.* model can be used by Portuguese district archives to speed up access to information, particularly in parish registers.

Keywords: district archives, Digital Palaeography, Digital Humanities, parish records, collective memory.

* Este trabalho é financiado por fundos nacionais através da FCT – Fundação para a Ciência e a Tecnologia, I.P., no âmbito do projeto UIDB/00311/2020 com o identificador DOI 10.54499/UIDB/00311/2020 (<https://doi.org/10.54499/UIDB/00311/2020>).

** Centro de História da Sociedade e da Cultura, Universidade de Coimbra, Portugal.
ORCID iD: <https://orcid.org/0000-0003-1247-8346>. E-mail: margaridadiasdasilva@gmail.com

Introdução

A memória coletiva tem existência material nos *monumentos*¹ e nos *documentos*. Os primeiros estão associados ao “poder de perpetuação, voluntária ou involuntária, das sociedades históricas (é um legado à memória coletiva) e o reenviar a testemunhos que só numa parcela mínima são testemunhos escritos” (Le Goff 1990, 536). Os segundos eram valorizados pela escola histórica positivista dos finais do século XIX e inícios do século XX, que, pela sua objetividade, pareciam “opor-se à intencionalidade do monumento”, afirmando-se, essencialmente, como testemunhos escritos (Le Goff 1990, 536).

Na verdade, o *documento*, que triunfa sobre o *monumento*, é fundamentalmente texto.

A escrita constitui-se como uma revolução tecnológica que permite a comunicação assíncrona, no espaço e no tempo, porque fixa a memória dos indivíduos e das sociedades. Uma revolução tão importante que marca a passagem da pré-história para a história.

A escrita pode definir-se como o cruzamento do “trinómio ‘cérebro vista e mão’” (Santos 1993, 552) e “como forma inigualável de comunicação, de acumulação de conhecimentos, de constituição de uma memória colectiva” (Santos 2000, 74).

O documento escrito surge associado à memória inscrita num qualquer suporte: “depois de tentativas sobre osso, estofos, pele, como na Rússia antiga; folhas de palmeira, como na Índia; carapaça de tartaruga, como na China; e finalmente papiro, pergaminho e papel” (Le Goff 1990, 433), tendo usado também a pedra, tabuinhas de argila e o couro. Os documentos escritos aparecem como consequência da atividade humana, e são “sempre um produto das condições sociais, económicas, políticas e culturais das sociedades” (Santos 2004, 13).

A escrita tem a dupla função de armazenar informações que podem ser comunicadas através do tempo e do espaço, pois “fornece ao homem um processo de marcação, memorização e registo”, e a possibilidade de “reexaminar, reordenar, retificar frases e até palavras isoladas” ao passar da esfera auditiva à visual (Le Goff 1990, 433).

A escrita registada constitui-se como objeto material de extrema importância para a civilização, pois a base do desenvolvimento da civilização humana é a ideia, e o avanço da alfabetização, da cultura, e da democracia só foram possíveis através da materialização da informação. A existência de um suporte, que vincula o(s) pensamento(s), permite ao mesmo tempo conhecer as alterações cronológicas relativas a negócios, educação, história, ciência, direito e artes, e legar conhecimento às gerações futuras (Gejman 2009, 389).

¹ Jacques Le Goff explica que: “A palavra latina *monumentum* remete para a raiz indo-européia *men*, que exprime uma das funções essenciais do espírito (*mens*), a memória (*memini*). O verbo *monere* significa ‘fazer recordar’, de onde ‘avisar’, ‘iluminar’, ‘instruir’. O *monumentum* é um sinal do passado. Atendendo às suas origens filológicas, o monumento é tudo aquilo que pode evocar o passado, perpetuar a recordação, por exemplo, os atos escritos” (Le Goff 1990, 535).

No entanto, com o triunfo do *documento* (texto escrito) sobre o *monumento* (legado da memória coletiva), os registos da acção humana ficaram restritos a uma elite que dominava a escrita e o escrito. A invenção da escrita foi, sem dúvida, algo revolucionário no processo de comunicação. A transmissão de mensagens através dos escritos foi, durante milénios, uma forma de comunicação exclusiva de grupos minoritários, pois,

enquanto a troca de informação cara a cara estava ao alcance de todos, os sistemas mais modernos, utilizados para transportar informação para além dos confins de uma família ou de uma aldeia, eram essencialmente fechados e utilizados para efeitos de controlo social ou político. Eram, com efeito, armas de elite. (Toffler 2000, 37)

Portugal, do século XVI ao século XVIII, continuou “a ser um país da oralidade, da memória auditiva, das tradições e dos costumes”, onde, no entanto, “a cultura escrita e o livro em particular foram ganhando peso, invertendo perspectivas antigas, impondo-se em novos modelos e práticas, mesmo se a maioria continuava a não saber ler e escrever” (Lisboa e Miranda 2011, 335).

No século XIX, em Portugal, as taxas de iliteracia superavam os 75% ou até os 80% (Lisboa e Miranda 2011, 339).

Na verdade, podemos considerar que até à universalização da escrita e da alfabetização generalizada, que na realidade portuguesa só acontecem depois da revolução democrática do 25 de abril de 1974, a escrita e o escrito permaneceram restritos a uma elite, a um grupo de privilegiados, a um grupo dominante, fosse ele político, religioso ou económico.

Em paralelo, desde sempre as administrações sentiram a necessidade de ler e transcrever documentos antigos, o que levou à especialização de profissionais numa técnica designada por “paleografia de leitura”, que ao longo dos séculos se manteve de forma consistente, fazendo jus à etimologia do conceito (Borges e Silva 2018, 3).

De paleografia empírica, no sentido restrito de prática de decifrar escritas/documentos antigos (Nunes 1973, 9), a ciência auxiliar, primeiro da Diplomática, depois da História, é só no século XX que a paleografia (pese embora a etimologia da palavra) vê renovados o seu objeto de estudo e método, atingindo assim a estatuto de ciência autónoma (Borges e Silva 2018, 35).

O século XXI trouxe novos usos e uma nova vida à ciência paleográfica, aproximando-a de um público mais vasto, sabendo aproveitar as potencialidades da Internet e da Web 2.0. Vejam-se, por exemplo, os inúmeros cursos de *e-learning* de ensino e prática da paleografia de leitura em arquivos europeus (Borges e Silva 2018) ou o crescente número de projetos de *crowdsourcing* e participação cidadã de transcrição massiva de documentos, que permitem que qualquer pessoa com acesso à Internet participe na transcrição e leitura de manuscritos (Borges e Silva 2018).

Na verdade, a prática da paleografia de leitura tem acompanhado a evolução das novas tecnologias, atualmente também fazendo uso da inteligência artificial (IA).

O trabalho em apreço apresenta-se como um estudo exploratório que procura, na relação entre a função de comunicação e acesso aos documentos e à informação com a Paleografia Digital, um método de trabalho nos arquivos distritais portugueses que acelere o acesso à informação, nomeadamente com a utilização do modelo de transcrição automática: *Portuguese handwriting 16th-19th c.*

Para cumprir o objetivo, o artigo subdivide-se em três pontos principais: primeiramente, a relação entre arquivos enquanto mediadores da informação e a evolução das tecnologias de informação e comunicação (TIC), com ênfase nos conceitos de acesso e comunicação; o segundo ponto é dedicado aos arquivos distritais, às suas funções e competências; e, por fim, explora-se o modelo *Portuguese handwriting 16th-19th c.* como método de acesso à informação dos registos paroquiais salvaguardados nos arquivos distritais portugueses.

Arquivos e mediação em ambiente digital

Os arquivos, as bibliotecas e os museus foram “uma criação da Modernidade inscrita numa concepção iluminista, revolucionária, liberal e, mais tarde, democrática, de partilha colectiva da Memória e da Cultura de um Povo e da Humanidade” (Silva 2006, 79)².

Com a Revolução Francesa (1789), emerge a ideia do Estado-Nação e a criação da instituição pública ou serviço cultural “Arquivo / Biblioteca / Museu Nacional” enquanto serviços de uma “Administração zeladora e promotora da “cultura superior ou do espírito”, entre outras diversas obrigações políticas no contexto do Antigo Regime, criados como “espaços próprios e singulares” que recriavam um passado (re)descoberto, conservando, guardando ou custodiando, “simultaneamente, princípio e fim de um modo de ver o Mundo mediatizado por fragmentos ou vestígios de um passado”, assentes na concepção do “Bem Comum” e na ideia de (re)construção da Memória nacional (Silva 2005, 29-30).

Em conjunto, arquivos, bibliotecas, centros de documentação e museus permanecem os grandes responsáveis pela preservação da maioria da memória registada do mundo, fontes para a construção e interpretação do conhecimento sobre as sociedades e os indivíduos no seu contexto; sendo que os arquivos são responsáveis, também, pelo fortalecimento das identidades locais ou nacionais, assim como pela prossecução de direitos e garantias dos cidadãos e das comunidades.

² Sobre a história das bibliotecas, arquivos e museus, cf., por exemplo, Docampo Capilla 2010, 60; López de Prado 2003, 6-10; Ortega 2004, 3; Silva et al. 2002, 46; Silva 2006, 73-79.

Com a invenção do computador na primeira metade do século XX, a memória eletrônica surgia como ilimitada, quando comparada à memória humana, que, embora milenarmente auxiliada pela existência de arquivos, bibliotecas, museus e dicionários, ganhava no computador um aliado que congregaria “toda a memória do mundo” (Morin 1977, 215).

Se a humanidade tem vivido sempre com base em “procesos informativos aunque no por cierto con la forma mediática actual de la información digital” (Capurro 2014, 112), a eletrônica trouxe consigo uma das maiores revoluções da história, cujas enormes consequências ainda estão por compreender na totalidade (Gejman 2009, 389).

No domínio da arquivística, a discussão em torno da mudança de paradigma nos arquivos em relação direta com os desenvolvimentos tecnológicos e a criação de documentos eletrônicos foi referida pela primeira vez por Hugh Taylor em 1987, na conferência de abertura do congresso anual da Associação dos Arquivistas Canadianos, intitulada: “Transformation in the Archives: Technological Adjustment or Paradigm Shift?” (Taylor 1987).

A partir da segunda metade do século XX, com a invenção e comercialização dos computadores, a criação e vulgarização da Internet e dos meios de comunicação social (*media*) ou de massas (*mass media*), e a digitalização massiva, assiste-se a uma explosão digital. Por sua vez, o novo milénio ficou marcado pela consolidação da Internet como veículo de comunicação de massa, pelo nascimento e rápido desenvolvimento das redes sociais (a primeira foi o Facebook em 2004), o lançamento da página do Youtube (em 2005), a popularização dos dispositivos móveis (*smartphones* e *tablets*), e, atualmente, com os desafios causados pela inteligência artificial.

É hoje consensual que as Tecnologias da Informação e da Comunicação (TIC), ampliadas pela Internet e a World Wide Web (WWW), e toda a revolução tecnológica associada, trouxeram um novo conjunto de questões, problemas e desafios à comunidade global dos profissionais da documentação e da informação. A forma de acesso e divulgação da informação alterou-se nos últimos anos, com grande preponderância para o digital, as plataformas da Web 2.0 ou as redes sociais a obrigarem as instituições de memória a desenvolver capacidades de adequação à cibercultura e ao acompanhamento das evoluções tecnológicas (Silva 2013).

Na verdade, as transformações da sociedade digital provocaram alterações na forma de comunicação em arquivos, bibliotecas, centros de documentação e museus, mas essas mudanças não foram unicamente potenciadas pelo digital e pelas tecnologias, estas instituições exploram e abraçam novos papéis para além da mera coleta e disseminação da informação (Audunson, Hobohm e Tóth 2020, 165).

No desenvolvimento das suas atividades, as instituições de memória adquirem, procesam, transformam, valorizam, transmitem, distribuem e disseminam informação para criar conhecimento e responder à satisfação das necessidades da sociedade.

Segundo o *Dicionário de terminologia arquivística*, cabe aos arquivos assegurar a “aquisição, conservação, organização e comunicação” (Alves et al. 1993, 7) da documentação/informação que salvaguardam. A comunicação é “função primordial do arquivo, que consiste em facultar aos utilizadores, atuais ou potenciais, informações, referências e documentos de que disponha e sobre os quais não recaia qualquer restrição de comunicabilidade” (Alves et al. 1993, 23). Por sua vez, para a Norma Portuguesa 4041, a comunicação é a função de “difundir o conhecimento do seu acervo documental e promover a sua divulgação” (NP 4041 2005, 19).

O acesso à informação é função derradeira (porque depende do cumprimento das que lhe antecedem) e relevante porque vai permitir recuperar informação pertinente aos investigadores e aos cidadãos em geral.

A Norma Portuguesa 4041, destinada a “fixar termos e conceitos de uso corrente na teoria e prática arquivística”, no ponto 4.7 “Funções dos serviços de arquivo: Comunicação”, define-se “acessibilidade” como a “disponibilidade dos documentos para consulta, em consequência da sua comunicabilidade e do necessário tratamento arquivístico (descrição arquivística, estado de conservação, etc.)” (NP 4041 2005, 19). Os arquivos permitem o acesso a documentos/informação que são “para consulta em consequência da sua comunicabilidade e existência de instrumentos de descrição documental” (Alves et al. 1993, 3).

O acesso à informação arquivística faz-se através de guias de fundos, inventários, catálogos e índices (genericamente designados por Instrumentos de Descrição Documental (IDD) ou instrumentos de recuperação da informação), hoje espelhados em plataformas de descrição arquivística elaboradas segundo as normas de descrição arquivística internacionais (ISAD(G)) – Archeevo, Digtarq ou AtoM (para citar as mais usadas no contexto arquivístico português).

As alterações tecnológicas ocorridas ao longo dos séculos revolucionaram a forma de comunicar dos seres humanos entre si e em sociedade.

Na verdade, as formas de comunicação atuais em arquivos passam pela inclusão dos utilizadores no processo infocomunicacional. Essa alteração da relação dos serviços com os utilizadores verifica-se, por exemplo, com a adoção das ferramentas da Web 2.0 e as folksonomias, no convite feito à participação e colaboração de todos em funções tradicionalmente desempenhadas exclusivamente pelos técnicos, nomeadamente, a descrição de conteúdos e a criação de pontos de acesso. De facto, a Sociedade da Informação veio alterar as formas de acesso à informação. O aparecimento da World Wide Web revolucionou a forma de relacionamento dos utilizadores, clientes e cidadãos com a informação. O desenvolvimento de ferramentas colaborativas no que se designa por Web 2.0 veio contribuir, de forma decisiva, para a construção colaborativa do conhecimento, fazendo uso

da inteligência coletiva, em que todos são simultaneamente consumidores e produtores de informação (Silva 2013).

A maior parte da documentação custodiada em arquivos é descrita ao nível da série, pois o grau de exaustividade e especificidade necessários para a descrição ao nível da peça não se coadunam nem com as capacidades humanas nem com as possibilidades financeiras das instituições, pela morosidade que isso implicaria. Isso provoca uma quantidade significativa de massa documental não tratada e é neste ponto que a Inteligência Artificial pode ser uma ferramenta útil aos arquivos e aos arquivistas.

Para alguns autores, a verdadeira revolução não é tecnológica, mas a mudança de atitude por parte dos profissionais da informação (Silva, Borges e Marques 2018; Audunson, Hobohm e Tóth 2020).

De facto, assiste-se ainda a uma “divisão digital” no mundo das Humanidades, entre os tradicionalistas que preferem o analógico e os “humanistas digitais” (Spina 2023, 115).

O avanço tecnológico que a IA representa neste domínio provoca receio no universo arquivístico, muito embora, desde 2018, a Câmara Municipal de Abrantes utilize a IA aplicada à gestão documental.³

O desenvolvimento de programas de reconhecimento óptico de caracteres (OCR) para material impresso foi momento marcante nas Humanidades Digitais, hoje de utilização fácil e corrente. Desde os anos 1990, com base em imagens digitais e IA, foram iniciados projetos de reconhecimento óptico de manuscritos com o objetivo de remover “the manuscript from the matrix of its paper model” (Stutzmann 2011, 219). Em 2013, a mesma cientista anunciou o triunfo da paleografia digital, apesar da complexidade da tarefa e do muito que ainda estava por conseguir (Stutzmann 2017).

Arquivos distritais portugueses: funções e competências

Portugal tem 18 arquivos distritais em território continental (o Arquivo da Universidade de Coimbra e o Arquivo Distrital de Braga, por razões históricas, estão dependentes das universidades de Coimbra e do Minho, respetivamente) e dois arquivos regionais nas ilhas da Madeira e dos Açores.

Após a instauração da República em Portugal, foram adotadas medidas de descentralização do grande volume de documentação à espera de dar entrada no Arquivo Nacional da Torre do Tombo. Júlio Dantas estava à frente da Inspeção das Bibliotecas e Arquivos

³ <https://www.computerworld.com.pt/2022/06/24/projeto-de-inteligencia-artificial-aplicada-a-gestao-documental-do-municipio-de-abrantes-desperta-interesse-nacional/>.

e foi nessa altura que foram sendo fundados os arquivos distritais, a partir de 1916, num processo longo e difícil, só concluído nos anos 80 do século XX (Carvalho 2014, 131-132; Janeiro 2020, 1-4; Ribeiro 1993, 30, 351, 599-601 e 615-616).

Cabe, hoje, à Direcção-Geral do Livro dos Arquivos e das Bibliotecas (DGLAB) (sucessor dos Arquivos Nacionais / Torre do Tombo fundados em 1992, por extinção do Instituto Português de Arquivos, criado pelo Decreto-Lei nº 152/88, de 29 de abril) a tutela dos arquivos distritais e regionais portugueses.

Os arquivos distritais salvaguardam a herança arquivística do respetivo distrito, produzidos no âmbito da Administração Pública, e incorporam obrigatoriamente: os livros dos registos paroquiais, documentação das conservatórias do registo civil, das conservatórias do registo do notariado, a documentação dos tribunais, e de todos organismos cessantes ou entidades extintas de proveniência diversa (públicos e privados, de entidades coletivas e singulares, religiosas e associativas, pessoais e familiares) da área geográfica do distrito, determinados pela lei. Por outras palavras, os arquivos distritais surgem como serviços de incorporação e salvaguarda da área geográfica correspondente à sede de distrito.

A Arquivística, ciência aplicada da Ciência da Informação, está hoje em diálogo e em relação com disciplinas tão diversas como: a Diplomática, a Paleografia, a História, a Administração, o Direito, as Ciências da Computação ou Informática (e as tecnologias em geral), a Gestão, entre outras. De forma breve, na evolução do processo de mediação em arquivos, o *arquivista-paleógrafo* e a Arquivística enquanto ciência auxiliar da História, marcados pela formação da *École des Chartres* e dos cursos em Ciências Documentais, evoluem para um novo paradigma resultante do desenvolvimento das TIC, onde o arquivista é um mediador, cientista/gestor da informação, promotor de dinâmicas colaborativas entre os consumidores e produtores de informação (que podem ser o mesmo) (Ribeiro 2011, 59-73; Silva e Ribeiro 2008, 79).

O papel de mediação dos arquivos distritais reside na sua capacidade de permitir o acesso aos seus arquivos e documentos, através da sua organização e da representação da informação em si contida de acordo com a normalização nacional e internacional. Assim, os arquivos distritais cumprem as funções de aquisição, preservação, organização, descrição, transformação, (re)utilização, recuperação, comunicação, divulgação e acesso a documentos/informação que são evidências sociais, intelectuais, artísticas e até espirituais da realidade territorial em que se insere.

As funções de acesso e de comunicação são fundamentais no processo de mediação realizado pelos arquivos distritais.

A presença *on-line* dos arquivos distritais portugueses permite-lhes ganhar visibilidade, num processo que se pretende de melhoria contínua da qualidade do acesso à

informação, que é considerando prioritário, “pois só o acesso público justifica e legitima a custódia e a preservação” (Silva 2006, 21).

Como os arquivos distritais portugueses estão sob tutela da DGLAB, as suas páginas de Internet são em tudo idênticas, seguem o mesmo *web design*, exceção feita às páginas do Arquivo da Universidade de Coimbra e do Arquivo Distrital do Minho, ambos dependentes das respetivas universidades (Silva e Borges 2022).

A transcrição automática nos arquivos distritais portugueses

Após o 25 de Abril de 1974 e o início do regime democrático em Portugal, os arquivos distritais portugueses iniciaram a informatização dos seus catálogos com a utilização de plataformas de descrição arquivísticas *on-line*.

Com os novos meios de comunicação, e uma informação “infinitamente crescente, rápida e complexa”, surge o problema de “como apreendê-la para melhor dominar sua produção, comunicação e uso, uma vez que as técnicas que a produzem, memorizam e veiculam atingem todos os dias os limites do infinitamente pequeno e do infinitamente grande?” (Le Coadic 2007, 238).

Se é verdade que a “automatização veio, afinal de contas, valorizar os instrumentos de acesso, graças às múltiplas possibilidades de pesquisa que os sistemas informáticos encerram” (Ribeiro 2003, 694), o aumento da disponibilização de conteúdos digitais *on-line* nem sempre tem significado o equivalente acesso à informação.

De facto, verifica-se que a transferência de instrumentos de descrição documental e de recuperação da informação para ambiente digital não tem sido acompanhada de maior exaustividade de descrição.

Veja-se o caso dos registos paroquiais dos diversos arquivos distritais portugueses.

A regra oficial e obrigatória de registar os batismos e os casamentos nas igrejas em livro próprio data do século XVI, mais precisamente do Concílio de Trento (1545-1563). Mais tarde, em 1614, o Papa Paulo V estabeleceu a mesma obrigação para o registo dos óbitos.

Estas disposições oficiais do Vaticano não eram, no entanto, uma novidade na maioria dos países católicos, nomeadamente em Portugal, que no primeiro quartel do século XVI registava os batismos e casamentos nos chamados livros mistos.

No século XX, o regime republicano promulgou a Lei de 20 de fevereiro de 1911, que criou o registo civil obrigatório em Portugal.

A partir de 1979, os arquivos distritais portugueses iniciaram uma política de digitalização. Principalmente devido aos protocolos assinados com a Igreja de Jesus Cristo dos

Santos dos Últimos Dias (ou Mórmones), os registos paroquiais (livros mistos, de batismos, casamentos e óbitos) têm sido das séries documentais mais disponibilizadas em linha.

Em pesquisa realizada em junho de 2023, os 18 arquivos distritais portugueses apresentavam os seguintes números de imagens relativas a registos paroquiais (**Fig.1**):

Arquivo Distrital	Objetos digitais de registos paroquiais	Datas
Aveiro	6217	séculos XVI-XIX
Beja	25 815	1550-1911
Braga	48 853	1583-1911
Bragança	44 935	?
Castelo Branco	10 616	1714-1890
Coimbra	20 899	1459-1911
Évora	20 729	1535-1911
Faro	10 640	1587-1911
Guarda	72 252	1530-1910
Leiria	10 952	1534-1911
Lisboa	71 191	?
Portalegre	4987	1533-1892
Porto	40 421	1536-1911
Santarém	13 096	1597-1911
Setúbal	10 927	1555-1911
Viana do Castelo	29 730	?
Vila Real	28 354	1533-1911
Viseu	36 477	1523-1911

1 Número de objetos digitais de registos paroquiais por arquivos distrital. Elaboração pela própria autora (recolha realizada em junho de 2023). Fonte: páginas web dos arquivos distritais portugueses.

A soma perfaz 507 091 de imagens. No entanto, a mera disponibilização de objetos digitais, sem informação arquivística associada ao nível do registo, não permite uma pesquisa rápida e eficaz. Significa isto que, atualmente, não é possível pesquisar termos (palavras ou texto) nos registos paroquiais.

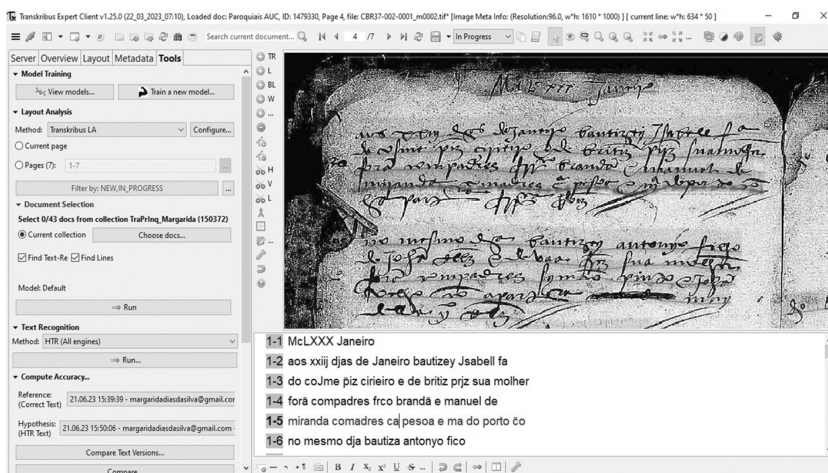
Se a comunicação é “o processo intermediário que permite a troca de informações entre pessoas” (Le Coadic 1996, 13), de facto, comunicar pressupõe um intercâmbio para a satisfação de uma necessidade, que envolve um emissor e um recetor, ambos conhecedores dos mesmos signos e símbolos que compõem a mensagem a transmitir.

Na Europa ocidental, de uma forma geral, houve lugar a “grandes empresas de cópias de cartas avulsas reunidas posteriormente em códice” (Santos 1993, 551) desde os séculos VIII-IX, e, em Portugal esta tendência surge já documentada desde o século XII, no que se tornou uma *praxis* nas chancelarias régias e monásticas (Cruz 1966; Marques 1981; Santos 1993). Em Portugal, destaca-se o monumental projeto da Leitura Nova, levado a cabo entre 1504-1552.

A Paleografia Digital é um campo fértil para auxiliar os arquivos e arquivistas na captação de informação, nomeadamente através do Reconhecimento de Texto Manuscrito (HTR - *Handwritten Text Recognition*), mas tem pouca ou nenhuma utilização em Portugal.

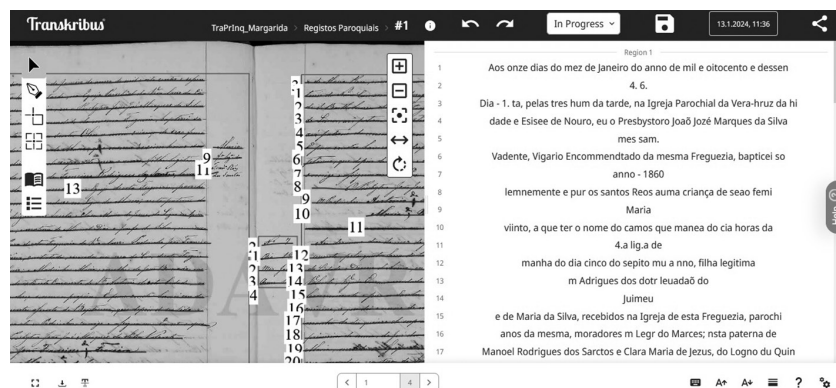
A plataforma Transkribus foi criada por investigadores da universidade de Innsbruck, na Áustria, e utiliza o HTR em conjunto com recursos de inteligência artificial para reconhecer “a morfologia das letras, as características linguísticas de cada língua, bem como o *modus scribendi* das mãos ali representadas e faça, com celeridade, transcrições automatizadas” (Lose et al. 2024, 9).

Numa experiência utilizando imagens de registos paroquiais portugueses no Transkribus, com a aplicação do modelo *Portuguese handwriting 16th-19th c.*, tornado público em setembro de 2023, verifica-se uma boa leitura da grafia do século XVI (**Fig. 2**).



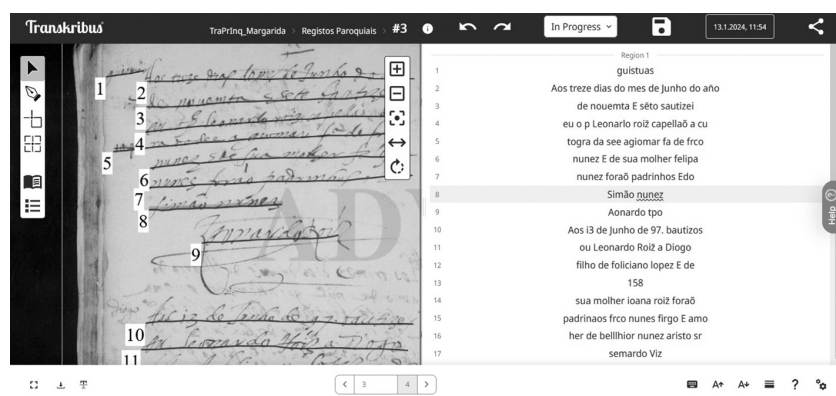
2 Livro de Batismos, paróquia de São Tiago, Coimbra, 1530 (CBR37-002-0001_m0002), Arquivo da Universidade de Coimbra. URL: <https://pesquisa.auc.uc.pt/details?id=35116>.

Também a grafia do século XIX é bem lida pelo modelo, no entanto, como se verifica pela **Fig. 3**, é sempre necessária a correção por parte do arquivista.



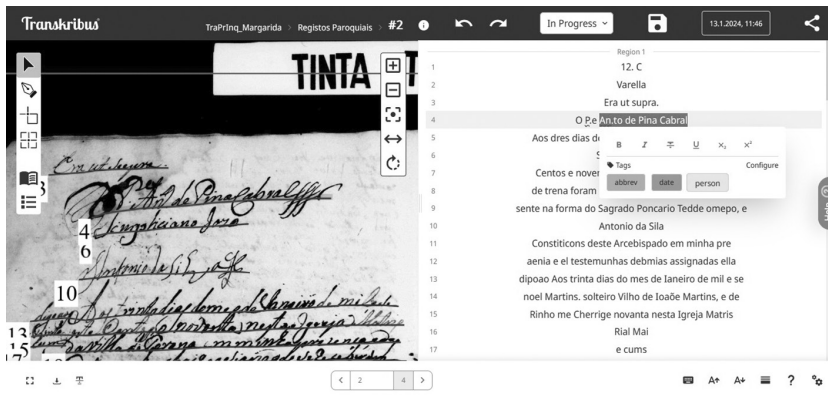
3 Livro de Batismos, paróquia de Vera Cruz, Aveiro, 1860 (PT-ADAVR-DIO-PAVR12-001-0001_m0006_derivada), Arquivo Distrital de Aveiro. URL: <https://digitarq.adavr.arquivos.pt/viewer?id=1293342>.

Apesar do modelo *Portuguese handwriting 16th-19th c.* ter sido criado a partir dos processos da Inquisição Portuguesa e o conteúdo ser diferente daquele dos registos paroquiais, há constantes que facilitam a utilização do modelo: estarem datados do mesmo período temporal (séculos XVI a XIX), conterem nomes portugueses idênticos (quer nomes próprios, quer apelidos) e localidades e geografias de Portugal (**Fig. 4**).

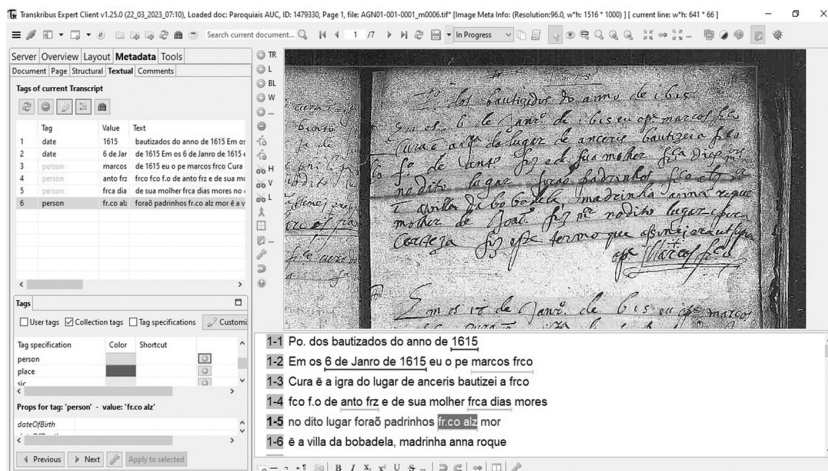


4 Livro de baptismos, paróquia de Ocidental, Viseu, (PT-ADVIS-PRQ-PVIS31-001-0001_m0022), Arquivo Distrital de Viseu. <https://digitarq.advis.arquivos.pt/viewer?id=1204071>

Outra grande vantagem do Transkribus, é a possibilidade de introdução de metadados e de colocar etiquetas (*tags*), como Pessoa, Local ou Data, por exemplo, e, ainda, de acrescentar propriedades (no caso da Pessoa, a Idade, o Pseudónimo, entre outros) (fig. 5, 6 e 7).



5 Livro de casamentos, paróquia de Santiago Maior de Terena, Évora, (PT-ADEV-PRQEV-ADL05-002-0001_m0012), Arquivo da Distrital de Évora. URL: <https://digitalq.adev.arquivos.pt/viewer?id=995696>.



6 Livro Mistos, paróquia de Anseriz, Arganil, (AGN01-001-0001_m0006), Arquivo da Universidade de Coimbra. URL: <https://pesquisa.auc.uc.pt/details?id=31763>.

Transkribus Expert Client v1.25.0 (22_03_2023_0710). Loaded doc: Paroquiais AUC, ID: 1479330, Page 5, file: CBR37-002-0001_m0181.tif [Image Meta Info: (Resolution:96.0, w/h: 1596 * 1000)] [current line: w/h: 598 * 117]

Server | Overview | Layout | **Metadata** | Tools

Document | Page | Structural | Textual | Comments

Tags of current Transcript

Tag	Value	Text
1	date	fo dois dias de janeiro digo quatro
2	date	de janeiro digo quatro dias tro bag
3	PERSON	Anto dias tres baggeas Anto escrauo de r
4	PERSON	mandel Anto escrauo de mandel Roiz e fo
5	PERSON	ma ser mandel Roiz e fo de ma serena ses
6	PERSON	ester fo padrinhos ester frz serada e a
7	PERSON	serada E a madrinha brites Roiz mu
8	PERSON	Jeronim Roiz molher de Jeronimo Roiz ce c
9	date	quinze Jeronimo Roiz ce quinze des janir

Tags

☐ User tags ☒ Collection tags ☐ Tag specifications ☒ Custom

Tag specification

Color

Shortcut

blackening

date

day

Props for tag: no tag selected

Previous Next Apply to selected

1-4 e fo de ma serena seu escrauo e
1-5 forã padrinhos ester frz serada
1-6 E a madrinha brites Roiz molher
1-7 de Jeronimo Roiz
1-8 pie quinze des Janiro de 508 baptis
1-9 cou hũ capelao do su bpo abel

7 Livro de Batismos, paróquia de São Tiago, Coimbra, (CBR37-002-0001_m0181), Arquivo da Universidade de Coimbra. URL: <https://pesquisa.auc.uc.pt/details?id=35115>.

Os dados extraídos são facilmente exportados para Word ou PDF que depois podem ser, por sua vez, exportados para as plataformas de descrição arquivística *on-line*.

Este processo de transcrição automática acelera o acesso à informação, embora não descure a existência de profissionais com conhecimentos de leitura paleográfica, sobretudo para corrigir a máquina.

É evidente que é impossível digitalizar e transcrever todos os documentos salvaguardados nos arquivos distritais. “However, it is also true that Artificial Intelligence like Transkribus and the creation of increasingly performing transcription models could boost the creation of a digital heritage that can guarantee concrete access to the sources of our Past” (Spina 2023, 118).

O grande volume de objetos digitais de registos paroquiais a transcrever não se coaduna com os recursos humanos existentes nos arquivos distritais, em geral. O que propo-nho complementarmente à utilização do modelo *Portuguese handwriting 16th-19th c.* é a promoção de projetos de transcrição colaborativa em torno da transcrição automática dos registos paroquiais. Conforme estudado, “a comunidade dos genealogistas desempenha papel importante na colaboração destes projetos” (Silva e Borges 2018, 5). Na verdade, este grupo muito específico, o dos genealogistas, enquanto utilizadores “egocêntricos”, querem saber mais sobre as suas origens e a sua família. Para aceder à informação relativa aos seus antepassados, cada utilizador contribuirá com essa identificação e a informação ficará acessível para toda a comunidade de investigadores e cidadãos em geral.

Conclusões

A democratização do acesso à informação tem sido conseguida através da disponibilização de conteúdos *on-line*, acessíveis 24 horas por dia, todos os dias, a quem tenha acesso a um computador com Internet.

Se o paradigma do século XX foi o da digitalização massiva de arquivos e coleções, no século XXI deseja-se o acesso aos dados. Disponibilizar objetivos digitais em imagens significa apenas uma transferência de suporte e não um verdadeiro acesso à informação contida nos documentos.

A crescente utilização de plataformas colaborativas da Web 2.0 ou de IA para a transcrição massiva de documentos têm dado 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.

Do nosso ponto de vista, a transcrição automática pode acelerar o acesso à informação e desbloquear milhares de dados, no entanto, esta não dispensa recursos humanos especializados, conhecedores e habilitados para a leitura paleográfica, nem recursos financeiros, para armazenamento dos dados e para utilização de plataformas de transcrição automática, como o caso do Transkribus. No entanto, é a partir da utilização do modelo e das correções que forem sendo efetuadas que a “máquina aprende” e o modelo se aperfeiçoa.

Do trabalho realizado conclui-se que a utilização do modelo *Portuguese handwriting 16th-19th c.*:

- é adequada aos registos paroquiais existentes nos arquivos distritais portugueses;
- acelera a transcrição dos documentos a partir das imagens;
- não dispensa conhecimentos de paleografia para correção de erros de leitura;
- não dispensa a alocação de recursos humanos e financeiros.

O grande volume de objetos digitais a transcrever beneficiaria, e muito, da promoção de projetos de ciência cidadã ou *crowdsourcing*.

Do que fica exposto, os arquivos distritais portugueses podem beneficiar do modelo HTR específico criado, em 2023, para grafias portuguesas dos séculos XVI a XIX: *Portuguese handwriting 16th-19th c.*

Bibliografia

- ALVES, Ivone. 1993. *Dicionário de terminologia arquivística*. Lisboa: Instituto da Biblioteca Nacional e do Livro.
- AUDUNDON, Ragnar, Hans-Christoph Hobohm, e M. Tóth. 2020. "LAM professionals and the public sphere". In *Libraries, archives and museums as democratic spaces in a digital age*, ed. Ragnar Audunson, Herbjørn Andresen, Cicilie Fagerlid, Erik Henningsen, Hans-Christoph Hobohm, Henrik Jochumsen, Håkon Larsen e Tonje Vold, 165-184. Berlim: De Gruyter Saur. <https://doi.org/10.1515/9783110636628>.
- BORGES, Leonor Calvão, e Ana Margarida Dias da Silva. 2018. "Transcrições em linha: e-learning de Paleografia em arquivos europeus". *Revista Portuguesa de História* XLVIII: 39-59. https://doi.org/10.14195/0870-4147_49_2.
- CAPURRO, Rafael. 2014. "Pasado, presente y futuro de la noción de información". *Logeion: Filosofia da Informação* 1 (1): 110-136. <https://doi.org/10.21728/logcion.2014v1n1.p110-136>.
- CARVALHO, Milena. 2014. "Estudo da mediação e do uso da informação nos Arquivos Distritais." Tese de Doutoramento, Faculdade de Letras da Universidade de Coimbra. <https://hdl.handle.net/10316/25994>.
- CONFRARIA, João, coord. 2002. *As Comunicações na Idade Contemporânea: cartas, telégrafo e telefones*. Lisboa: Fundação Portuguesa das Comunicações.
- CRUZ, António. 1966. "Observações sobre o estudo da paleografia em Portugal". *Cale: Revista da Faculdade de Letras do Porto* 1: 173-233.
- DOCAMPO CAPILLA, Javier. 2010. "Bibliotecas de museos: panorama internacional de una tipología bibliotecária". *Educación y Biblioteca* 22 (176): 60-71.
- GEJMAN, Robert. 2009. "An integrated framework for information, communication and knowledge definitions". *tripleC* 7 (2): 386-398. <https://doi.org/10.31269/triplec.v7i2.103>.
- JANEIRO, Jorge. 2020. "Processo de criação dos arquivos distritais: síntese histórica". *Cadernos BAD* 1-2: 1-10. <https://doi.org/10.48798/cadernosbad.2489>.
- LE COADIC, Yves-François. 1996. *A ciência da informação*. Brasília: Briquet de Lemos.
- LE COADIC, Yves-François. 2007. "A Matemática da Informação". In *Para entender a ciência da informação*, org. L.M.B.B. Toutain, 219-239. Salvador: EDUFBA.
- LE GOFF, Jacques. 1990. *História e memória (Coleção Repertórios)*. Campinas: Editora da UNICAMP.
- LISBOA, João Luís, e Tiago C. P. dos Reis Miranda. 2011. "A Cultura escrita nos espaços privados". In *História da Vida Privada em Portugal*, org. José Mattoso. Vol. 2: *A Idade Moderna*, coord. Nuno Gonçalo Monteiro, 334-394. Lisboa: Círculo de Leitores; Temas e Debates.

- LÓPEZ DE PRADO, María del Rosário. 2003. "Bibliotecas de museos en España: Características específicas y análisis DAFO". *Revista General de Información y Documentación* 13 (1): 5-35. <http://revistas.ucm.es/index.php/RGID/article/view/RGID0303120005A>.
- LOSE, Alicia Duha, João Guilherme Andrade dos Santos, Leonardo Coelho Marques de Jesus, Livia Borges Souza Magalhães, e Lucia Werneck Xavier. 2024. "Transkribus: uma ferramenta de paleografia digital mediando pesquisas em fontes inquisitoriais". *Revista LaborHistórico* 10 (1): e63285. <https://doi.org/10.24206/lh.v10i1.63285>.
- MARQUES, Oliveira. 1981. "Paleografia". In *Dicionário de História de Portugal*, org. Joel Serrão. Vol. 4, 529. Porto: Livraria Figueirinhas.
- MORIN, Edgar. 1977. *O Método*. Vol. 1: *A Natureza da Natureza*. Mem Martins: Publicações Europa-América.
- MORIN, Edgar. 1996. *O Método*. Vol. 3: *Conhecimento do conhecimento*. Mem Martins: Publicações Europa-América.
- NORMA PORTUGUESA 4041. 2005. *Informação e documentação terminologia arquivística: conceitos básicos*. Lisboa: Instituto Português da Qualidade.
- NUNES, Eduardo Borges. 1973. "O conceito novo de paleografia". *Portugaliae Historica* 1: 9-12.
- ORTEGA, Cristina Dotta. 2004. "Relações históricas entre biblioteconomia, documentação e ciência da informação". *DataGramaZero: Revista de Ciência da Informação* 5 (5): artigo 3.
- RIBEIRO, Fernanda. 2003. "O sistema de informação arquivística da Universidade do Porto: Potenciar o uso da memória informacional retro-prospectivamente". *Páginas a&b* 11: 79-99.
- RIBEIRO, Fernanda. 2011. "A Arquivística como disciplina aplicada no campo da ciência da informação". *Perspectivas em gestão & conhecimento* 1 (1): 59-73.
- SANTOS, Maria José Azevedo. 1993. "Uma ciência em Portugal e na Europa: A paleografia (séculos XIX-XX)". *Theologica* 28 (2): 551-567.
- SANTOS, Maria José Azevedo. 2000. *Ler e compreender a escrita na Idade Média*. Lisboa: Colibri.
- SANTOS, Maria José Azevedo. 2004. *Assina quem sabe e lê quem pode*. Coimbra: Imprensa da Universidade de Coimbra.
- SILVA, Ana Margarida Dias da. 2013. "O uso da Internet e da Web 2.0 na difusão e acesso à informação arquivística: O caso dos arquivos municipais portugueses". Dissertação de mestrado, Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa.
- SILVA, Ana Margarida Dias da, e Leonor Calvão Borges. 2018. "A transcrição e a leitura de manuscritos entre o crowdsourcing e a participação cidadã". *Atas 13º Congresso Nacional BAD*. <https://publicacoes.bad.pt/revistas/index.php/congressosbad/article/view/1792>.

- SILVA, Ana Margarida Dias da, e Leonor Calvão Borges. 2022. "Evaluating User Experience of the Portuguese District Archives websites". *Qualitative and Quantitative Methods in Libraries* 3 (11): 587-601.
- SILVA, Ana Margarida Dias da, Leonor Calvão Borges, e Maria Beatriz Marques. 2018. "Crowdsourcing in history projects in local archives of Portugal and England", In *International Conference on Information Society and Smart Cities (ISC)*, dir. Hani Hamdan. Cambridge: Cambridge University. <https://hdl.handle.net/10316/81355>.
- SILVA, Armando Malheiro da. 2005. "Informação, cultura e património: Uma abordagem exploratória feita no campo emergente da ciência da informação". In *Conservar para quê? 8.ª Mesa-Redonda de Primavera*, coord. V. O. Jorge, 27-58. Porto: Faculdade de Letras da Universidade do Porto.
- SILVA, Armando Malheiro da. 2006. "Arquivo, direito e poder: Para uma desconstrução construtiva". In *Archivum et Jus: Ciclo de Conferências*, ed. Maria José Azevedo Santos, 61-90. Coimbra: Arquivo da Universidade de Coimbra.
- SILVA, Armando Malheiro da, Fernanda Ribeiro, Júlio Ramos, e Manuel Luís Real. 2002. *Arquivística: Teoria e prática de uma ciência da informação*. Porto: Edições Afrontamento.
- SILVA, Armando Malheiro da, e Fernanda Ribeiro. 2008. *Das "ciências documentais" à ciência da informação. Ensaio epistemológico para um novo modelo curricular*. Porto: Edições Afrontamento.
- SPINA, Salvatore. 2023. "Handwritten Text Recognition as a digital perspective of Archival Science". *AIDAinformazioni* 1-2: 115-132.
- STUTZMANN, Elisabeth. 2011. "Nouvelles technologies au service de la codicologie et de la paléographie". *Scriptorium* 65 (1): 217-223.
- STUTZMANN, Elisabeth. 2017. "Paléographie, une révolution numérique", *L'Histoire* 439: 30-31.
- TAYLOR, Hugh A. 1987. "Transformation in the Archives: Technological Adjustment or Paradigm Shift?". *Archivaria* 2: 12-28. <https://archivaria.ca/index.php/archivaria/article/view/11451>.
- TOFFLER, Alvin. 2000. *A Terceira Vaga (Vida e Cultura)*. Lisboa: Livros do Brasil.