



Anabela Borges  
Teles Ribeiro

## Acervos portugueses *on-line*

Projetos de digitalização e disponibilização *on-line* realizados por instituição portuguesa contribuíram de forma decisiva para a preservação dos documentos originais de valor inestimável de um dos acervos documentais mais importantes daquele país, o Arquivo Nacional da Torre do Tombo.

> Os Arquivos Patrimoniais têm implementado medidas de digitalização e disponibilização *on-line* como forma de divulgar e preservar o seu património. A informatização dos arquivos é relevante no contexto de diversas atividades culturais, educativas, científicas, bem como em projetos inovadores e criativos. Salienta-se a sua relevância no contexto de globalização e a sua importância para a afirmação da identidade nacional.

Em 2003 e 2004 a DGARQ (ex-Instituto dos Arquivos Nacionais/Torre do Tombo) implementa a digitalização no âmbito do primeiro projeto de digitalização sistemática, designado *TT On-line*,<sup>1</sup> mas apenas para capturar imagens a partir de suportes intermédios (microfilmes e fotografia).

Entre 2005 e 2008 a DGARQ, seguindo as linhas orientadoras de medidas estratégicas de digitalização de âmbito nacional e europeu, dá início a uma política de digitalização sistemática e direta do seu património e para isso teve de reforçar as suas capacidades de execução interna em matéria de *hardware*, *software* e formação dos recursos humanos. Dá-se início a uma nova política de transferência de suportes e simultaneamente à sua preservação em formato digital, contribuindo com mais e melhores conteúdos *on-line* e assegurando a preservação a longo prazo dos conteúdos digitais.

O projeto *TT On-line* englobou uma coleção de documentos designada Tesouros da TT, composta por 229 documentos e 5.348 imagens, que abrange o período cronológico de 1211-1908; a Coleção do Corpo Cronológico, composta por 45 mil documentos e 296.719 imagens, que abrange o período cronológico de 1137-1699; a série Diários, do fundo Arquivo Oliveira Salazar, composta por 72 volumes, no total de 21.595 imagens, que abrange o período cronológico de 1933-1968; e o Arquivo Fotográfico da Companhia de Moçambique, composto por 7 mil fotografias, que abrange o período cronológico de 1888-1942. Esses

documentos foram selecionados por sua relevância para a comunidade, por se encontrarem em mau estado de conservação e alguns deles por serem pouco conhecidos, mas com potencial de interesse. O objetivo foi divulgá-los à comunidade de uma forma global e facilitar o acesso gratuito a alguns dos principais fundos e coleções documentais da Torre do Tombo, potencializando esses conteúdos em benefício da investigação, do conhecimento, do património, da cultura e, simultaneamente, adquirir infraestruturas tecnológicas – *hardware* e *software*.

Entre os diversos projetos de digitalização que se seguiram, realçam-se alguns deles pela importância de conteúdos, nomeadamente o projeto *Memórias paroquiais, 1758: uma Coleção on-line*,<sup>2</sup> coleção constituída por 44 volumes, também designada Dicionário Geográfico, com elevado índice de consulta e pedidos de reprodução; o projecto *Instrumentos de Descrição – on-line*,<sup>3</sup> que deu acesso nesse formato aos instrumentos de descrição documental antigos da Torre do Tombo, que na sua maioria continuavam a ser consultados apenas através do seu original, no Serviço de Referência. Nesse sentido, foram produzidos cerca de 40 mil registros de metainformação, bem como igual quantidade de imagens (matrizes e cópias) dos instrumentos de descrição. Para cada documento integrado na coleção de instrumentos de descrição da Torre do Tombo, foi ainda elaborada uma descrição à qual se associou a respectiva metainformação arquivística, destacando-se os termos de indexação iniciais e finais de cada fólio ou página/imagem digitalizada, de modo a facilitar a sua pesquisa. A solução encontrada veio permitir ainda a disponibilização na internet dos instrumentos de descrição na íntegra, podendo o usuário, de modo mais amigável que anteriormente, efetuar a navegação por meio de marcadores (*bookmarks*), produzidos a partir de termos de indexação recolhidos no momento da elaboração da metainformação arquivística.

O projeto *PIDE/Registo Geral de Presos (1932-1974)* abrangeu a digitalização de 148 livros, no total de 64.354 imagens que foram descritas no nível do documento simples, e 29.510 registros criados. As imagens foram renomeadas de acordo com a sua descrição, de forma a criarem-se termos de indexação de acordo com o número de registro de cada fólio ou página/imagem digitalizada, facilitando assim a sua pesquisa. As imagens matrizes foram capturadas diretamente dos originais, em escala de cinzentos, com uma resolução de 200 dpi, em formato TIFF, sem compressão, com metainformação técnica associada, permitindo a criação de um produto digital normalizado, com qualidade suficiente para garantir a sua autenticidade, tendo em vista a sua preservação a longo prazo. Desses volumes apenas um deles se encontra *on-line* e os restantes estão disponíveis apenas na intranet, mas com acesso restrito devido às reservas de comunicação, de acordo com o artigo 17º da Lei 16 de 23 de janeiro de 1993. Essa solução foi adotada a partir do Projecto ID-on pois permite facilmente efetuar a navegação por meio de marcadores (*bookmarks*), produzidos com base em termos de indexação recolhidos no momento da elaboração da metainformação arquivística.

O projeto da série *PIDE/Propaganda apreendida* contou com a descrição normalizada dos 665 documentos e disponibilização de 4.064 imagens de formato compreendido entre A5 e A0, em cores; os projetos de digitalização de *Documentação paroquial* desenvolvidos pelos serviços dependentes da DGARQ, em colaboração com a Family Search (ex-SGU), nomeadamente: projeto de digitalização do Arquivo Distrital do Porto, Portalegre, Faro, Vila Real e Aveiro e os projectos de digitalização a partir de microfilme de documentação paroquial desenvolvidos em colaboração com as autarquias de Torres Vedras, Odivelas, Grandola, Lisboa/Gabinete de Estudos Olisiponenses e com a Associação dos Amigos da Torre do Tombo, bem como os projetos de digitalização do Arquivo Distrital de Leiria e Setúbal, os quais, na sua globalidade, têm disponibilizado 2.476.054 de imagens.

No que respeita aos fundos e coleções fotográficas, têm sido desenvolvidos trabalhos de digitalização sistemática, nomeadamente a série *Reportagem Política*, sub-série *Política Geral*, do fundo Secretariado Nacional de Informação (SNI), entre 1970 e 1985, no total de 7.900 imagens; a série *Primeiros Ministros “Oliveira Salazar”*, do fundo SNI, num total de 1.820 imagens; a série *Álbuns Gerais*, subsérie *Álbuns Alfabéticos* do fundo *Empresa Pública Jornal o Século*, que se encontra disponível *on-line*, referentes aos álbuns que compreendem o período de 1926 a setembro de 1938, no total de 12.211 imagens. Os álbuns referentes ao período de 1938-1969 estão disponíveis em formato digital nas instalações do ANTT.

#### Acessibilidade em linha

Em 2007, a DGARQ desenvolveu uma estratégia para a acessibilidade em linha de seus acervos patrimoniais, em conformidade com as medidas da Agenda Europeia para a Cultura e medidas de âmbito nacional, contribuindo para incentivar a criatividade, a competitividade e o crescimento, tendo se beneficiado do apoio do Ministério da Cultura e da Rede Energética Nacional para a implementação *on-line* do Projecto *Inquisição de Lisboa*, prevendo a descrição, conservação, digitalização e disponibilização *on-line* de documentos,<sup>4</sup> iniciativa que teve como principal objetivo a disponibilização *on-line* dos documentos pertencentes ao acervo daquela instituição, existentes na Torre do Tombo.

Estão disponíveis *on-line* 19.775 registros descritivos e 2.320.576 imagens. Esse projeto contribuiu de forma decisiva para a preservação dos documentos originais de um dos arquivos mais importantes do Arquivo Nacional da Torre do Tombo, uma vez que a sua consulta passou a ser efetuada em ambiente digital.

A adoção da tecnologia digital nesses projetos implicou sempre a prévia descrição da documentação de acordo



A estrutura física dos processos e livros exigiram a utilização de equipamentos em sistema planetário, com dois pratos reguláveis na abertura e altura, com vidro que se pode recolher automaticamente, quando necessário, para virar as páginas do documento, com luz fria e modo de captura de imagem linear.

A captura de imagem de cada página foi isolada, com a utilização de folhas de cartolina preta, quando esta, devido à diversidade de formato ou ao mau estado de conservação (perfurações, perdas de suporte ou outras, como diversidade de formatos dos cadernos), interfere na informação da página anterior ou posterior. O tratamento de imagens executado foi feito apenas para ajustar a imagem digital à imagem real do documento original, ou seja, garantir que apenas a área útil seja representada, sem perda da integridade e autenticidade do original.

O espaço ocupado em servidor pelas imagens da Inquisição de Lisboa é de 80 *terabytes*, o que exigiu à DGARQ a renovação do seu sistema de armazenamento e segurança de informação, já que num curto espaço de tempo se exigiu um enorme crescimento, pois o volume de informação a gerir teve de se assentar em pilares sólidos que garantissem a preservação da informação em formato digital.

O produto final é um catálogo estruturado, que disponibiliza em linha registros descritivos com objetos digitais associados, capazes de serem preservados a longo prazo acessíveis ao público em geral via internet.

#### A digitalização e disponibilização *on-line* do património

A tecnologia digital permite uma enorme diversidade de aplicações e nesse sentido é necessário que as entidades detentoras de documentos definam bem os objetivos dos trabalhos, para que os especialistas nessa tecnologia possam ajudar a decidir quais as melhores

especificações técnicas a adotar. Recomenda-se aos arquivos definitivos que essas se assentem em procedimentos normalizados de preparação, descrição, captura, tratamento, gravação, controle e armazenamento para a produção de matrizes/objetos digitais, com capacidade de garantir a autenticidade e a integridade dos documentos originais, bem como a sua preservação e acessibilidade a longo prazo. As matrizes têm como funções primárias servirem de arquivo de conservação permanente e como repositório documental para a produção de derivadas, podendo ser tratadas como imagens originais, quando essas não existem, ou ser usadas como forma de segurança contra a perda de originais devido a desastres.

A disponibilização é realizada a partir das matrizes com elevada qualidade de imagem, sendo criadas imagens derivadas (cópias) produzidas com especificações técnicas que devem ser decididas em função das características dos documentos originais: desde formatos inferiores a A5 até formatos superiores a A0; grafia de pequena dimensão e não normalizada; suportes de diferentes cores; mau estado de conservação etc. e que visam garantir a qualidade da informação a ser colocada *on-line*. Assim, a velocidade de acesso das derivadas pode ser, por vezes, comprometida pelas características dos originais, devido ao seu peso em *kbyte*, mas em seu próprio benefício, porque dessa forma o usuário tem a possibilidade de aceder a documentos com resolução mais alta e em maior formato, o que lhe permite maior facilidade de interpretação. O usuário pode, no entanto, descarregar os documentos com qualidade de leitura e utilizar técnicas de processamento de forma a ajustá-los à medida das suas necessidades (compressão, redução de resolução etc.)

As matrizes podem ser processadas por meio de diferentes métodos e nesse sentido a produção de derivadas depende da aplicação usada em cada organização. Saliente-se que as entidades detentoras de documentação que pretendem desenvolver trabalhos ou projetos

de digitalização deverão realizar previamente a análise e caracterização da documentação – conhecer a especificidade e diversidade dos fundos, coleções, séries documentais etc., em termos de nível de descrição ou catalogação; estrutura física; cores do suporte e grafia; formato; estado de conservação e o nível de informatização. Nos projetos deverá englobar-se ainda a análise das próprias condições do sistema de informatização dos fundos/coleções, velocidade de rede, *hardware* e *software* existentes.

As especificações técnicas recomendadas são ditadas pela natureza dos documentos e pelo produto final pretendido, mas, depois de conhecidos todos os detalhes, concluímos que em arquivos definitivos, na maioria das vezes, o modo de captura de imagens que permite preservar os documentos é o planetário: mesas com prensa, com dois pratos reguláveis na abertura, com vidro amovível devido ao estado de conservação dos documentos; com luz fria; dimensão mesa/prensa: 610mm x 450mm x 170mm.

No que diz respeito ao formato de fichários, esse varia de acordo com a política adotada, mas sempre que o objetivo é o de preservação – e para a gravação de matrizes adotamos o formato TIFF, com compressão nível zero,<sup>6</sup> resolução mínima de 300 dpi<sup>7</sup> esquema de cor RGB<sup>8</sup> e profundidade de 24 bits.<sup>9</sup> Essas especificações são de relevante importância quando pretendemos criar repositórios de documentos, porque se estes aspectos não forem garantidos teremos de repetir consecutivamente os processos de transferência de suporte para satisfazer as diferentes necessidades dos usuários.

Já as imagens derivadas para acesso podem ser produzidas pela própria aplicação informática, de acordo com as definições de perfis definidos previamente, sendo as especificações, na generalidade, em formato JPG; compressão variável de acordo com as características dos originais e com uma resolução  $\geq 150$  dpi.



Procedimentos técnicos de identificação, preparação e digitalização de documentos pertencentes aos fundos existentes na Torre do Tombo. Fotografias de Anabela Ribeiro, 2010. Arquivo Nacional da Torre do Tombo – TSO/IL.

A captura de imagens com qualidade de arquivo requer o cumprimento de alguns procedimentos, no sentido de respeitar os princípios de produção e regras dos documentos analógicos, nomeadamente:

- a) A captura das imagens deve apresentar o documento na sua globalidade, encadernação e pormenores;
- b) A captura deve ser efetuada página a página e respeitar a ordem de paginação/foliação;
- c) A informação das páginas deve ser isolada quando, devido à diversidade de formato ou mau estado de conservação (perfurações, perdas de suporte ou outras, diversidade de formatos dos cadernos), essa interfere na informação do fólio anterior ou posterior;
- d) Os fólios em branco devem ser capturados se estiverem numerados e se não representarem mais do que cinco imagens. Mas, no caso de este número ser ultrapassado, devem ser capturados apenas o primeiro e o último, aplicando-se o mesmo procedimento de captura para conjuntos de fólios que se encontrem em branco e sem numeração. Essa situação deve ser reportada em campo próprio na descrição do documento, designada pela norma ISAD(G) “Dimensão”;
- e) Aplicação da escala de cor, em cada objeto digital, para se poder ajustar à cor do original.

No contexto digital, os fichários de imagem são representações dos documentos e, para serem geridos e preservados, requerem uma estratégia de identificação. Nesse sentido, eles são organizados e nomeados de acordo com as especificações previamente definidas pelo sistema da DGARQ. Adotaram-se as especificações definidas no documento *Designação dos ficheiros de imagens*, Versão 2, DGARQ, 2003. Assim, a organização de fichários se assenta numa estrutura hierarquizada, de acordo com os códigos de referência fornecidos pelo sistema de descrição normalizado com base na ISAD(G) para cada documento, permitindo a identificação e a recuperação dos objetos digitais.

Na captura das imagens matrizes é imprescindível a associação metainformação técnica normalizada, pelo que se recomenda no mínimo a produção de 34 elementos, devendo esses serem conservados com as próprias matrizes.

O tratamento de matrizes deve apenas ser realizado para ajustar a imagem digital às dimensões da imagem do documento original, ou seja, garantir que apenas a área útil seja representada, sem perder a integridade e a autenticidade do original. A matriz deve ser guardada antes que qualquer tipo de processamento seja realizado, ainda que esse venha a ser necessário por razões de acessibilidade à informação.

O controlo de qualidade deve ser efetuado desde a captura até o momento da disponibilização. Com esse controle, visa-se garantir a conformidade com os requisitos definidos para cada uma das etapas. Considera-se adequado que a sua realização corresponda a apenas a 10% do universo, com registro em fichas específicas, em que constem:

1. especificações técnicas detalhadas de cada objeto digital (OD), data e assinatura de responsáveis;
2. controle de qualidade arquivístico de cada OD.

O armazenamento digital recomendado é um sistema em rede, de longo prazo, desenvolvido exclusivamente para conteúdos fixos *Content Addressed Storage* (CAS). O sistema CAS introduz nova eficiência de energia, nova tecnologia de discos e soluções de gestão de arquivos permanentes. Esse visa guardar e disponibilizar grandes volumes de informação, facilitar o aumento da capacidade de armazenamento, garantir a alta performance e a proteção dos dados. Nesse sistema, a ingestão dos OD é realizada em lotes de imagens. Esses lotes só são considerados em condições de entrega quando se verifica o seguinte: os objetos digitais se encontram íntegros e normalizados e estejam produzidos

os registros de controle de qualidade definidos. O remetente das imagens deve conservar localmente a informação original, até o momento em que lhe seja comunicada a aceitação do lote, e só após a sua aceitação é que ele poderá eliminar esse lote localmente.

A disponibilização *on-line* dos documentos originais é efetuada de forma gratuita e com critérios de racionalidade no sentido de se encontrar um equilíbrio entre qualidade e peso das imagens/páginas dos documentos, no sentido de facilitar o acesso *web* aos usuários, sem contudo comprometer os documentos em pior estado que necessitam de mais qualidade.

Como forma de salvaguarda dos direitos da entidade detentora dos documentos, foi adotada uma marca óptica aplicável automaticamente pelo uso da informática, quando gera as imagens derivadas, a colocar *on-line*, a fim de identificar a entidade detentora dos documentos e assim garantir os direitos de propriedade. O usuário que necessitar de documentos sem essa marca óptica terá de contactar o detentor dos documentos e solicitar as imagens, justificando a sua necessidade.

### Considerações finais

Após alguns anos de trabalho com essa tecnologia, constatamos que as formas e metodologias de ação se alteraram e, para que se assegure a continuidade de disponibilização *on-line* de documentos, necessitamos reforçar a normalização das descrições, descrever com melhor qualidade e criar registros com os elementos mínimos exigidos para o intercâmbio de informação, sem comprometer a disponibilização.

*Hardware* e *software* existentes no mercado têm evoluído de acordo com as necessidades manifestadas pelas organizações. Existem hoje soluções tecnológicas de excelente qualidade dotadas de condições para respeitar

todas as exigências colocadas pela documentação. Contudo, em muitos casos, as organizações veem-se limitadas a adotar verdadeiras soluções de preservação digital a longo prazo, simplesmente por razões orçamentárias, e não por falta de tecnologia.

A DGARQ tem implementado diferentes projetos com equipes de especialistas, entre arquivistas, informáticos, técnicos de digitalização e de conservação, que têm desenvolvido conhecimentos capazes de possibilitar avanços significativos na aplicação de novos e melhores projetos de digitalização, preservação digital e disponibilização *on-line*.

Em nosso caso, o investimento realizado teve ainda impacto em diversos níveis, a saber:

- na adoção de novas formas de reprodução, com recurso à tecnologia digital;
- no início da disponibilização documental através do recurso da internet;
- na potencialização das descrições arquivísticas;
- na preservação dos documentos que se encontram *on-line*, via digital, permitindo um acesso integral ao respectivo conteúdo, sem os riscos de desgaste dos suportes originais.

Esse formato documental permite ainda o restauro virtual, garantindo assim a leitura de partes da informação quase ilegíveis nos próprios documentos. Por sua diversidade e abrangência de conteúdos, salienta-se também o impacto desses projetos em nível socioeconômico, uma vez que propiciam um acesso quase incondicional, em qualquer local e a qualquer hora, abrindo a todos os interessados suas fontes de informação e as imagens dos documentos associados relativos a alguns fundos e coleções de maior procura do Arquivo Nacional, ou de dilatado interesse potencial, como sucede no caso do Arquivo Fotográfico da Companhia de Moçambique e Grupo Entrepósito Comercial de Moçambique.

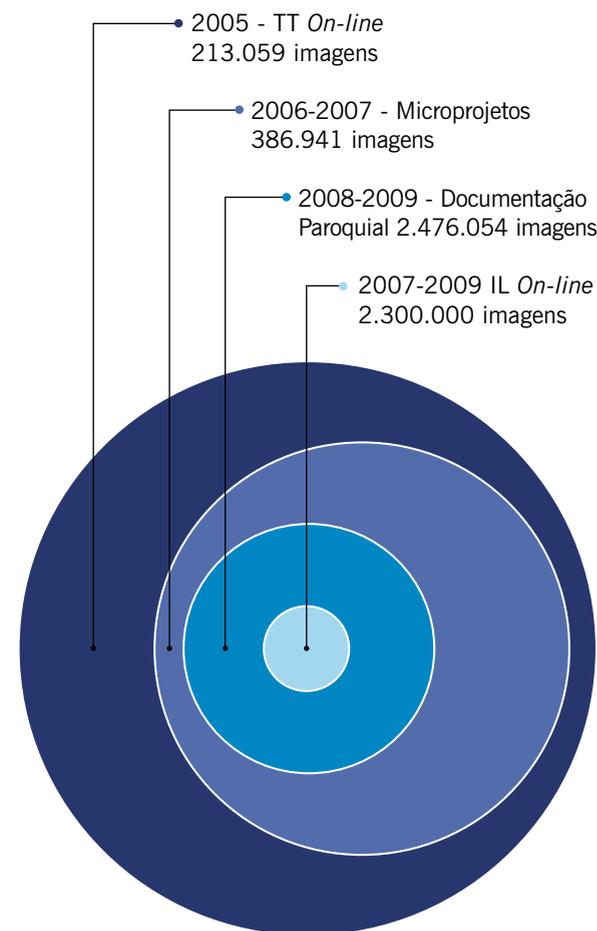
Esses projetos apresentam ainda, entre outras, as seguintes vantagens.

- a) Pesquisa mais rápida e mais eficaz, podendo-se fazer cruzamentos entre termos, de modo a obter resultados mais otimizados.
- b) Recuperação de informação que se encontra inacessível nos próprios originais, por meio do recurso a ferramentas de tratamento virtual.
- c) Disponibilização de descrições de documentos que estavam acessíveis aos usuários apenas por meio dos instrumentos de descrição tradicionais, em papel, existentes na Sala de Leitura e/ou publicados, os quais podem atualmente ser consultados por todos os cidadãos, de forma remota.
- d) Campo aberto para o intercâmbio com entidades nacionais – universidades, institutos ou outras entidades do setor científico e cultural –, bem como com entidades internacionais, no sentido de se estabelecerem protocolos de colaboração, uma vez que a documentação disponibilizada é muito abrangente.
- d) Em nível interno, uma otimização de recursos humanos, dado que os técnicos de arquivo deixaram de consultar os documentos originais para efetuar o seu trabalho, com maior rentabilidade, a partir dos documentos que se encontram digitalizados. Essa forma de trabalho permite ainda a rentabilização de outros recursos internos, ligados ao fornecimento dos documentos (trabalho da técnicas da Sala de Leitura e dos próprios depósitos).
- e) Estabelecimento de uma via tecnológica segura para o desenvolvimento de novos projetos que permitam a disponibilização de outros fundos documentais.

A disponibilização de conteúdos documentais, de forma remota, é uma mais-valia aos arquivos e à Torre do Tombo, em específico, uma vez que potencializa, dessa forma, o patrimônio arquivístico sob sua guarda, contribuindo assim para a necessária afirmação nacional. Trata-se de informatizar e digitalizar, de modo

sistemático, os arquivos históricos e o patrimônio cultural como meio para facilitar o seu tratamento e utilização, bem como para preservar e divulgar, por meio da internet, o nosso saber, acumulado ao longo de gerações.

Entre 2003 e 2009 a DGARQ disponibilizou *on-line* 5.366.054.000 imagens, como se demonstra a seguir.



#### Notas |

1. RIBEIRO, Anabela. TT *On-line*. *Boletim Arquivos Nacionais*, Lisboa, n. 12, p. 7, out.-dez. 2006.
2. BRAGA, Joana; RIBEIRO, Anabela; RUNA, Lucília. Memórias Paroquiais: uma Coleção Online – *Boletim Arquivos Nacionais*. Lisboa. ISSN 1645–5460. Nº 15 (Jan. - Mar. 2006), p.1-2.
3. PENTEADO, Pedro; RIBEIRO, Anabela. Novos caminhos da disponibilização da informação na Torre do Tombo: digitalização de instrumentos de descrição antigos. *Boletim Arquivos Nacionais*, Lisboa, n. 18, p.1-2, out.-dez. 2006; n. 15, p.1-2, jan.-mar. 2006.
4. RIBEIRO, Anabela; SERRO, Jorge; Ó RAMOS, Fátima; FIGUEIREDO, Catarina; TREMOCEIRO, Paulo. A Inquisição de Lisboa na nova plataforma digital: acesso e disponibilização. *Boletim Arquivos Nacionais*, Lisboa, n. 10, p. 5, jul.-set. 2009.
5. Disponível em: <http://dgarq.gov.pt/rede-portuguesa-de-arquivos/pesquisar-arquivos/catalogo/>.
6. ISO/TS 22028-2:2006 - Photography and graphic technology – Extended colour encodings for digital image storage, manipulation and interchange - Part 2: Reference output medium metric RGB colour image encoding (ROMM RGB).
7. ISO 3664:2000, *Viewing conditions—Graphic technology and photography*; ISO/DIS 16067-1, *Photography—Spatial resolution measurements of electronic scanners for photographic images—Part 1: Scanners for reflective media*; ISO/WD 16067-2, *Photography—Electronic scanners for photographic images—Spatial resolution measurements—Part 2: Film scanners*.
8. ISO/TS 22028-3:2006 - Photography and graphic technology – Extended colour encodings for digital image storage, manipulation and interchange - Part 3: Reference input medium metric RGB colour image encoding (RIMM RGB).
9. ISO/TS 22028-3:2006 - Photography and graphic technology – Extended colour encodings for digital image storage, manipulation and interchange - Part 3: Reference input medium metric RGB colour image encoding (RIMM RGB).
10. Z39.87. 2002, *Data Dictionary: Technical Metadata for Digital Still Images*. Draft Standard for Trial Use. NISO/AIIM. 61 p.
11. Por meio do site <http://digitarq.dgarq.gov.pt>

**Anabela Borges Teles Ribeiro** é licenciada em História pela Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa e especialista em Ciências Documentais – Opção Arquivo, pela Faculdade de Letras da Universidade de Lisboa. Doutoranda em Bibliografía y Documentación Retrospectiva en Humanidades, pela Universidad de Alcalá (DEA), onde obteve também o diploma de Estudos Avançados. Atualmente é chefe da Divisão de Gestão de Projetos da Direção Geral de Arquivo e do Núcleo de Reprodução do Arquivo Nacional da Torre do Tombo.

