

BIBLIOTECAS DIGITAIS: DIVULGAÇÃO E PRESERVAÇÃO

As bibliotecas digitais comportam valiosos benefícios. Mas a sua criação, além de exigir um cuidadoso planeamento, pressupõe a sua manutenção no tempo, a eficácia da recuperação dos dados, o incremento dos conteúdos e a atualização tecnológica.

Texto de Dália Guerreiro

[DOUTORANDA EM CIÊNCIAS DA INFORMAÇÃO E DA DOCUMENTAÇÃO PELA CIDEHUS: CIDEHUS-UE/FCT, UNIVERSIDADE DE ÉVORA, BOLSA DE SFRH / BD / 82229 / 2011]

BIBLIOTECA DIGITAL

Uma biblioteca digital pode definir-se como um conjunto de documentos digitais disponibilizados em linha, com mecanismos de pesquisa e de recuperação da informação. A disponibilização dos acervos através das bibliotecas digitais confere um elevado índice de visibilidade dos documentos, fomenta a preservação dos originais e permite o acesso a pessoas com necessidades especiais ou em determinadas circunstâncias de inacessibilidade.

Na 36.^a Conferência Geral¹, a UNESCO aprovou o *Manifesto de las IFLA/UNESCO sobre las Bibliotecas Digitales* | IFLA (2011). Neste texto, define-se que a biblioteca digital deve ser constituída de acordo com as normas internacionais, ser persistente no tempo e manter os conteúdos devidamente organizados e estruturados, usando as novas tecnologias no acesso e na recuperação da informação. “Digital libraries offer unique ways of recording, preserving, and propagating culture in multimedia form.” (WITTEN e BAINBRIDGE,

1. Realizada de 25 de outubro a 10 de novembro de 2011 em Paris.

2003, p. 5). Ou seja, uma biblioteca digital é um conjunto de documentos, imagens, sons, textos, vídeos - nados digitais ou digitalizados - colocados em linha na *web*, devidamente organizados e estruturados, usando as novas tecnologias no acesso e na recuperação da informação.

A criação e a divulgação de bibliotecas digitais incrementam a visibilidade das coleções (acesso/difusão) e permitem a consulta da mesma obra por vários utilizadores em simultâneo, preservando o original. Para o utilizador, é vantajoso o acesso direto às obras, em qualquer horário ou local. Além disso, é possível adequar o *layout* a indivíduos com incapacidades.

PLANEAMENTO DE COLEÇÕES DIGITAIS

A criação de uma coleção ou de uma biblioteca digital exige uma planificação antecipada. Ainda antes de pensar nas obras a digitalizar, ou no modo de o fazer, deve-se caracterizar-se a biblioteca/serviço, definir a sua missão e objetivos, identificar os recursos humanos disponíveis e respetivas competências, e fazer o levantamento do equipamento existente ou a adquirir. Ou seja, é necessário fazer uma caracterização, o mais completa possível, quer da biblioteca quer dos serviços e do pessoal envolvido. Isto implica ponderar se a instituição tem meios para prosseguir na elaboração de um projeto de digitalização, partindo do princípio de que este não vai ser uma tarefa excepcional mas mais uma rotina adicional a implementar (Cf. SITTS, 2000).

Uma vez analisada a biblioteca e confirmada a existência de condições políticas, económicas, materiais e de pessoal, pode avançar-se para a fase seguinte, que consiste na caracterização da biblioteca digital a criar. Depois de esboçar o plano, em traços largos, há que detalhar, proceder à programação minuciosa, no âmbito da qual se deve responder, de forma completa e concisa, às seguintes questões:

- Que trabalho tem de ser feito?
- Quem deverá fazê-lo?
- Onde deverá ser feito?
- Quando será feito?
- Como será feito?

A propedêutica do projeto é determinante para a eficácia da sua execução e para uma resposta adequada a uma audiência intrinsecamente diversificada em termos de competências e expectativas.

CRITÉRIOS DE SELEÇÃO

A definição dos critérios de seleção está intimamente ligada à missão da instituição e às características dos respetivos acervos da biblioteca. No entanto, podem referir-se algumas considerações genéricas que auxiliam na seleção das obras.

Preferencialmente, devem ser colocadas em linha obras que estejam em domínio público (art. 31.º do Decreto-Lei n.º 16/2008, de 1 de abril). O direito de autor caduca, na falta de disposição especial, 70 anos após a morte do criador intelectual. Se as obras estiverem protegidas pelo direito de autor, é necessário obter autorização escrita do autor, ou dos herdeiros ou seus representantes, e do editor para a disponibilização em linha.

DEVE CONSIDERAR-SE, COMO FATOR PREFERENCIAL, O VALOR PATRIMONIAL, HISTÓRICO, CULTURAL OU ARTÍSTICO DAS OBRAS, SEM PREJUÍZO DA COERÊNCIA DO CONJUNTO"

Deve considerar-se, como fator preferencial, o valor patrimonial, histórico, cultural ou artístico das obras, sem prejuízo da coerência do conjunto e em função do propósito global da biblioteca digital, tal como foi previamente definido.

Outro critério relevante é o da acessibilidade, que determina a criação de suportes alternativos para obras em reserva, de acesso restrito, ou num estado de conservação frágil. Pode, além disso, considerar-se a oportunidade de efetuar uma campanha de restauro simultânea à digitalização.

As dimensões do objeto são, também, um dado



PÁGINA PRINCIPAL
do portal do Registo
Nacional de Objetos
Digitais (RNOD)

importante na seleção. Obras de grandes dimensões, superiores a A0 (84,1 x 118,9 cm), são difíceis de digitalizar com os meios convencionais garantindo a obtenção de imagens com boa resolução, além de que a sua inclusão, em regra, aumenta exponencialmente os custos do projeto. Além destes, deve também avaliar-se os níveis de utilização da coleção, pelo que a identificação das obras mais solicitadas ou consultadas deve ser considerada prioritária.

Depois de elaborada a lista prévia de obras a digitalizar, é conveniente verificar se alguma já terá sido digitalizada e disponibilizada em linha por outras instituições. Em Portugal, para o efeito, pode consultar-se o Registo Nacional de Objetos Digitais² (RNOD), um agregador de conteúdos digitais disponibilizados por entidades portuguesas. Também deve ser consultada a EUROPEANA³, biblioteca virtual dos países da União Europeia e o Internet Archive⁴, entre outras bibliotecas digitais de referência.

DIGITALIZAÇÃO⁵

A digitalização, enquanto processo que permite a transformação do sinal analógico num sinal digital, é também designada por conversão A/D. O objeto analógico é representado digitalmente através de um conjunto discreto de amostras, cujo resultado se denomina representação digital (Cf.

RIBEIRO, 2004, pp. 31-39). O principal objetivo da digitalização de documentos com valor patrimonial é a obtenção de uma matriz de qualidade, dado que, independentemente do fim a que se destina, este processo deve garantir a preservação dos originais.

As imagens digitais obtêm-se a partir da digitalização dos originais, utilizando digitalizadores (vulgo *scanners*) ou máquinas fotográficas digitais. Em ambos os casos, as imagens resultantes são matriciais.

Existem vários modelos de representação das cores para fazer a conversão para o digital, sendo os mais utilizados: o RGB (Red, Green, Blue) e o CMYK (Cyan, Magenta, Yellow, Black). Em qualquer dos modelos, porém, há que considerar que a visão humana consegue distinguir uma gama de cores superior à dos equipamentos, pelo que a cor exibida em ecrã pode não ser exatamente igual à do original.

O número de *bits* utilizado para codificar, ou descrever um pixel designa-se por profundidade de cor, ou número de *bits* por pixel (bpp). A resolução de uma imagem digital corresponde ao produto do comprimento pela largura em pixels. A expressão pixel por polegada (ppp) refere-se à densidade do objeto, ou seja, à relação entre o número de pixels da imagem e a dimensão do original em polegadas, isto é, dividindo a altura e a largura em pixels pela dimensão da imagem em polegadas.

Existem vários formatos para processar os dados consoante a sua tipologia. Isto significa que os dados são codificados de acordo com normas fixas para cada um dos formatos existentes, entre os quais o formato TIFF. A capacidade para guardar imagens não comprimidas e sem perdas de dados, bem como a respetiva compatibilidade com as diversas plataformas (*iOS, Windows, Linux, etc.*), habilitam o TIFF a constituir-se como um formato de imagem adequado à realização de cópias digitais para preservação. Em contrapartida, para disponibilizar imagens

2. Disponível em <http://rnod.bnportugal.pt/rnod/>.

3. Disponível em <http://www.europeana.eu/>.

4. Disponível em <https://archive.org/index.php>.

5. Adaptado de GUERREIRO, Dália Maria Godinho - *Repositório digital de património cultural móvel: uma aplicação a objectos do culto católico* [Dissertação de Mestrado apresentada no ISCTE, Lisboa. 2009].

em linha, tem vindo a dar-se preferência ao formato JPEG, dado que este permite definir vários índices de compressão, sendo que o tamanho dos ficheiros diminui com o aumento do grau de compressão e numa relação direta com o índice de perda de qualidade da imagem.

Tendo como ponto de partida imagens com detalhe, com tons de cor contínuos e 24 bits de cor, considera-se:

Imagens de arquivo – formato TIFF (sem interpolação, nem compressão): preferencialmente coloridas, sem compressão. Estes ficheiros, arquivados como matrizes, são utilizados para gerar as cópias de consulta.

Imagens de consulta – formato JPEG: imagens de baixa resolução, mas com qualidade suficiente para disponibilizar via *web*, garantindo a legibilidade dos respetivos conteúdos.

As imagens-matriz devem incluir as cunhas de cor e de cinza, adequadas às dimensões do original para garantir a qualidade da respetiva preservação (Vd. GUERREIRO, 2009, pp. 41-43).

EQUIPAMENTO DE DIGITALIZAÇÃO

Os equipamentos utilizados para a aquisição de uma imagem digital bidimensional e fixa agrupam-se em duas grandes categorias: digitalizadores e câmaras digitais. A distinção entre ambos consiste na distância de focagem. Enquanto numa câmara digital a luz chega ao sensor de captação de imagem CCD através de um conjunto de lentes, permitindo a captação de imagens de objetos que se encontrem a diferentes distâncias, nos digitalizadores a distância do motivo a captar ao sensor é fixa, permitido apenas pequenas variações na profundidade de campo. A escolha do equipamento de digitalização para cada projeto depende dos seguintes fatores:

- tipologia de originais a digitalizar, de forma a minimizar os danos provocados pela digitalização;
- dimensões dos originais a digitalizar;
- resolução ótica máxima do equipamento;
- profundidade de cor;
- gama dinâmica;
- velocidade de digitalização.



No âmbito restrito da construção de repositórios digitais de espécies patrimoniais, podemos distinguir duas tipologias principais de digitalizadores: os de mesa e os planetários.

Os digitalizadores de mesa são preferenciais para digitalizar material opaco de pequena espessura como, por exemplo, estampas, fotografias, folhas soltas, herbários, moedas, medalhas, amostras de tecidos, habitualmente objetos de pequena dimensão e de pouca espessura, assim como de material transparente e diapositivos. Não são aconselhados para a digitalização de espécies encadernadas.

Os digitalizadores planetários consistem numa mesa articulada com compensação de lombada. Os equipamentos deste tipo, atualmente disponíveis no mercado, permitem uma resolução até 800 ppp óticos para uma área de digitalização até A0. Estes equipamentos, especialmente concebidos para a digitalização de livros, baseiam o seu funcionamento no comportamento inerente à leitura: o livro é colocado sobre um tampo articulado, o que permite a compensação da lombada, sujeitando as folhas com um vidro transparente de cor neutra e cuja pressão é calibrada por sensores.

MAPA
com as cunhas.
BNF, Ge D 26179 Rés

/// A DIGITALIZAÇÃO INTEGRAL DE OBRAS ENCADERNADAS COM VALOR PATRIMONIAL - E, COM MAIORIA DE RAZÃO, AS QUE EXIJAM REQUISITOS ESPECIAIS DE MANUSEAMENTO E DE PRESERVAÇÃO, COMO É GENERICAMENTE O CASO DO LIVRO ANTIGO - DEVE SER REALIZADA EM SISTEMAS PLANETÁRIOS"

DIGITALIZAÇÃO DE LIVROS

A diversidade de espécies em papel existentes numa biblioteca é vasta. A digitalização de coleções bibliográficas comporta riscos no âmbito da preservação, os quais derivam sobretudo do manuseamento e da sujeição a tensões físicas e a circunstâncias de iluminação invulgares. No caso das espécies encadernadas, crescem os problemas inerentes à abertura num ângulo suficiente para permitir a captação integral do conteúdo da obra, ocasionando eventuais danos na encadernação e, sobretudo, na lombada.

Os livros que tenham sido muito manuseados, ou cuja encadernação não seja muito apertada, permitem, em geral, um grau de abertura a 180°, considerada adequada para a digitalização em equipamentos planetários, mantendo a obra sujeita sob um vidro com filtros ultravioleta. Em contrapartida, as obras sem manuseamento ou submetidas a encadernações posteriores, sobretudo resultantes de restauros recentes, com a costura dos cadernos demasiado apertada e lombadas sem flexibilidade, implicam a utilização de um berço que mantenha a obra com uma abertura entre 100° e 90°.

Outro problema relacionado com a encadernação prende-se com as obras demasiado aparadas, adaptando a dimensão do livro à uniformização da coleção e ao tamanho das estantes, deixando, com isso, margens demasiado pequenas e,

mesmo, suprimindo alguma informação, como as notas marginais ou partes impressas. Para comprovar que essas lacunas constam do exemplar copiado, e como princípio, durante a digitalização deve colocar-se uma folha negra ou contrastante sob cada folha a digitalizar, deixando uma margem tão pequena quanto possível, não excedendo em 3% as dimensões do original. A digitalização sem esta margem adicional não permite discernir se a eliminação ou o corte de elementos perceptíveis na imagem derivam do original aparado ou de erros na digitalização.

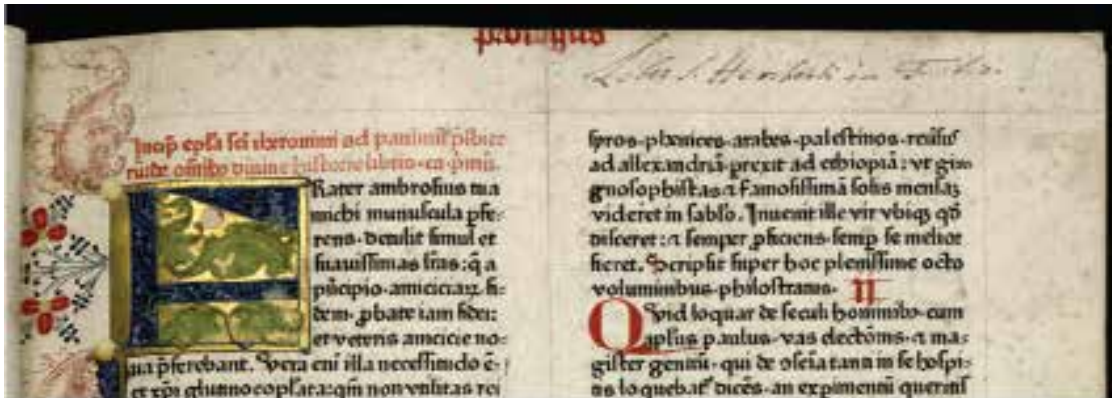
Em regra, a digitalização é feita na íntegra, ou seja, de capa a capa. Sobretudo nos casos em que o livro conserva a encadernação original, esta tem um valor patrimonial, pelo que também deve ser preservada através da transferência de suporte.

Por princípio, a digitalização deve fazer-se no sentido da leitura, o que implica uma correta observação da obra, sobretudo quando se trata de livros em línguas não latinas (como, por exemplo, o hebraico ou o árabe), averiguando se aquela se faz da esquerda para a direita ou da direita para a esquerda e certificando-se da posição correta dos caracteres, dado o risco de digitalizar uma obra de forma invertida.

A digitalização, por norma, deve fazer-se a 100% (escala 1:1), respeitando as dimensões do original. Nos digitalizadores de mesa ou nos planetários, esta definição surge por defeito; nas máquinas fotográficas digitais, é necessário introduzir este enunciado no início de cada sessão. Para a digitalização de preservação e para a conversão em formato texto, as imagens devem ser captadas a 300 ppp ou 600 ppp e em formato TIFF, sem interpolação, nem compressão, incluindo as cunhas de cor e de cinzentos. Desta forma, pode garantir-se que a obra é submetida apenas uma vez ao processo de digitalização, sendo a partir da matriz que – através da edição de imagem – se criam outras versões, nomeadamente para a colocação em linha, com resoluções mais baixas, noutros formatos e eliminando as cunhas de cor, mantendo o original sem alteração.

O plano de captação da imagem deve incluir a cunha de cor e a cunha de cinzentos, colocadas

IMAGEM
digitalizada com e sem
margem



tangencialmente à margem inferior do original, tendo em vista o posterior controlo de qualidade e o tratamento das imagens.

A digitalização integral de obras encadernadas com valor patrimonial – e, com maioria de razão, as que exijam requisitos especiais de manuseamento e de preservação, como é genericamente o caso do livro antigo – deve ser realizada em sistemas planetários, excluindo-se liminarmente os digitalizadores de mesa e de rolo e as máquinas fotográficas digitais do segmento doméstico.

ORGANIZAÇÃO DAS IMAGENS

Independentemente da tipologia dos originais, uma das tarefas integrantes e essenciais do processo de digitalização é a correta nomeação das imagens e das respetivas pastas. Nesse sentido, torna-se adequada a criação de um sistema de pastas e subpastas, correspondendo estas a cada uma das versões de imagens criadas (matriz em TIFF e exemplares de consulta em JPEG de diversas resoluções, PDF, etc.).

O nome da pasta, que deve identificar univocamente o exemplar físico digitalizado, passa a constituir o <identificador> da obra digitalizada. É a cota do objeto, correspondendo ao número de inventário na instituição, podendo apresentar a sigla da instituição antes do respetivo identificativo. O nome das subpastas regista o identificador e as características da imagem, de acordo com o esquema:

- <identificador>
- <identificador>_TIF
- <identificador>_TIF_t0
- <identificador>_JPG
- <identificador>_JPG_24-C-R0120
- <identificador>_JPG_08-G-R0150
- <identificador>_JPG_24-C-W0140

A nomeação das imagens faz-se de acordo com o seguinte esquema:

identificador_número sequencial_número da página_caraterísticas.formato

AS FICHAS DE CATALOGAÇÃO OU INVENTÁRIO, QUE ORIENTAVAM O REGISTO DA INFORMAÇÃO ELEMENTAR, VARIAVAM DE INSTITUIÇÃO PARA INSTITUIÇÃO

Nestas fórmulas, cada campo de informação é separado por traço baixo () e os vários elementos dentro de cada campo são separados por hífen (-).

No caso dos livros, o número sequencial da digitalização deve apresentar quatro dígitos, permitindo uma numeração de 0001 a 9999, sendo esta sequência reiniciada sempre que se inicia a digitalização de um novo volume, mesmo que da mesma obra.

O número de página reflete a paginação do original. No caso de obras com foliação (numeração sequencial dos retos), a indicação dos versos faz-se indicando o número do respetivo fólio seguido pela letra v (verso). Quando as partes preliminares ou finais da obra não sejam paginadas ou foliadas, pode criar-se uma numeração árabe ou romana, segundo um princípio de alternância em relação ao corpo da obra e de acordo com a respetiva catalogação. Isto é, se o corpo da obra apresentar uma paginação com numeração árabe, as restantes partes são numeradas a romano, ou vice-versa.

Nas obras não paginadas ou com paginação por assinatura é criada uma paginação árabe sequencial. O mesmo acontece nas obras com múltiplos erros de paginação (mais de três), salvaguardando-se o facto de haver discrepâncias entre o número de página do original e da cópia digital, através de uma nota inserida no respetivo registo de exemplar.

METADADOS

Metadados – ou informação acerca da informação – é um termo que designa, genericamente, todo o conjunto de dados aplicados à identificação de um determinado recurso. “Metadata is data about data. The term refers to any data used

to aid the identification, description and location of networked electronic resources.” (IFLA, 2005).

Embora aplicado em particular no âmbito das novas tecnologias da informação, o conceito não lhes é exclusivo, dado que, sempre que se organiza um acervo bibliográfico, gera-se um corpo de dados coerente que informa acerca do documento escrito para efeitos da sua descrição catalográfica, inventário ou indexação. Todas estas práticas sempre se assumiram como atividades programáticas das instituições detentoras de espólios.

A investigação parte da observação direta da peça e da recolha dos dados que lhe são essenciais (título ou designação, autoria, data e local de produção, matéria e forma). Para além destes, pode abranger questões mais alargadas, nas quais se inclui o estabelecimento de relações com outras peças congéneres e a identificação de dados que esclareçam acerca dos respetivos contextos originais.

Todos estes procedimentos são transversais às atividades biblioteconómica e a toda a documentação em torno do património móvel. Contudo, os critérios que presidiam ao registo da informação pautavam-se pela diversidade, na maioria das vezes sem critérios normativos. As fichas de catalogação ou inventário, que orientavam o registo da informação elementar, variavam de instituição para instituição. Neste aspeto, a introdução das novas tecnologias e, nomeadamente, o aparecimento das bases de dados informatizadas como estrutura de suporte à inventariação, impôs a normalização na organização e registo da informação.

Os metadados apresentam todas as vantagens da gestão de informação que é própria dos sistemas de bases de dados e acrescentam-lhes a interoperabilidade entre sistemas. Dessa forma, basta formular a questão num motor de busca para que este, através da metainformação, encontre a resposta em vários sistemas distintos. “Metadata is machine understandable information for the Web” (W3C, 2001). Essa informação destina-se essencialmente a ser lida por máquinas, permitindo a localização da informação no mundo digital.

Além disso, a informação tende a tornar-se mais abrangente, englobando dados relacionados com a versão digital do documento ou do objeto: “os metadados necessários para utilizar e gerir com sucesso objetos digitais são diferentes e mais vastos que os metadados utilizados para gerir coleções de obras impressas e outros materiais físicos” (LC, 2006).

Essa informação integra, por um lado, os dados identificativos do recurso, a sua descrição física, formal e material ou a compilação do respetivo conteúdo e, por outro, a referência aos intervenientes, programas e equipamentos utilizados na criação do produto digital ou digitalizado e a definição dos parâmetros de utilização ou condições de acesso. Os metadados fornecem respostas às questões de “quem”, “o quê”, “quando”, “onde”, “como” e “porquê”, formuladas sobre as várias facetas da informação que se pretende documentar. Constituem, deste modo, o somatório da informação disponível sobre determinado recurso: o conteúdo, o contexto e a respetiva estrutura. Este conjunto de dados abre novas perspetivas à investigação, na medida em que a pesquisa pode ser significativamente facilitada através de metadados ricos e consistentes (Cf. GILL, 1998, p. 8).

De entre os vários esquemas de metadados destaca-se o Dublin Core (DC). O esquema de metadados como o DC (COVER, 2005) é constituído por um conjunto de elementos (*elements set*), aos quais se associam as respetivas etiquetas. Por sua vez, o conjunto de elementos possui qualificadores (especificações do elemento), utiliza vocabulários controlados específicos, como os tesouros, e tem uma semântica e uma sintaxe próprias. A semântica, ou o elenco do significado dos elementos, enumera as várias etiquetas e os respetivos significados; a sintaxe, ou o conjunto de elementos bem definidos e as suas relações, permite hierarquizar as várias etiquetas e definir a forma como se vão relacionar umas com as outras. Cada elemento pode ter o seu esquema de codificação.

CONCLUSÃO

A construção de uma biblioteca digital tem por objetivo a constituição de um corpus coerente de informação, tendo em vista a preservação, o estudo, a comunicação e a divulgação dos materiais que lhes estão na origem. Isto coincide, em larga medida, com as funções patrimoniais atribuídas às bibliotecas. Porém, na atualidade, estas funções dispõem de ferramentas tecnológicas que lhes conferem uma nova dimensão, permitindo a transferência dos acervos para suporte digital e a divulgação, a nível global, através da *web*.⁹

BIBLIOGRAFIA

- COVER, Robin – *Technology reports: Metadata Encoding and Transmission Standard (METS)*. Cover pages: online resource for markup languages technologies [online]. 2005. [Consult. 16 abril 2007]. Disponível em <http://xml.coverpages.org/mets.html>.
- GILL, Tony et al. – *Introduction to Metadata: pathways to digital information* [online]. Ed. Murtha Baca. Los Angeles: Getty Research Institute, 1998, atual. 2008. [Consult. 4 janeiro 2009]. Disponível em http://www.getty.edu/research/conducting_research/standards/intrometadata/index.html.
- GUERREIRO, Dália Maria Godinho – *Repositório digital de património cultural móvel: uma aplicação a objectos do culto católico*. Lisboa: ISCTE, 2009. Dissertação de mestrado. Disponível em <http://hdl.handle.net/10071/1829>.
- IFLA – *Manifiesto de las IFLA/UNESCO sobre las bibliotecas digitales* [online]. Atual. 2011. [Consult. 18 junho 2014]. Disponível em <http://www.ifla.org/ES/publications/manifiesto-de-las-ifla-unesco-sobre-las-bibliotecas-digitales>.
- RIBEIRO, Nuno Jorge Gonçalves de Magalhães – *Multi-média e Tecnologias Interactivas*. Lisboa: FCA, 2004.
- SITTS, Maxine (ed.) – *Handbook for digital projects: a management tool for preservation and access*. Andover Massachusetts: Northeast Document Conservation Center, 2000.
- WITTEN, Ian H.; BAINBRIDGE, David – *How to build a digital library*. San Francisco: Morgan Kaufmann Publishers, 2003.