

# Bibliotecas, arquivos e outras coisas digitais

*José Borbinha*

INESC-ID – Grupo de Sistemas de Informação  
Rua Alves Redol 9, Apartado 13069 – 1000-029 Lisboa  
Tel: 213100317  
Email: jlb@ist.utl.pt

## RESUMO

Este artigo apresenta uma descrição dos principais conceitos a tomar em consideração na abordagem ao tema das bibliotecas digitais. Apesar do termo usado poder transmitir uma perspectiva apenas centrada na biblioteca, pretende-se demonstrar que o contexto e os conceitos em causa podem e devem ser da mesma forma alargados aos arquivos e museus. O artigo inicia-se com uma resenha histórica dos principais momentos e factos que contribuíram para a relevância do paradigma digital nestes meios. É apresentado de seguida um diagnóstico do momento e expectativas actuais para o futuro próximo, incluindo uma descrição dos principais desafios na agenda actual. Nestas análises são levadas em consideração argumentações considerando igualmente factores tecnológicos, organizacionais e políticos. Dada a longa lista de temas, por falta de espaço os mesmos são intencionalmente apresentados de forma simples e minimalista, visando antes de tudo proporcionar ao leitor não definições ou descrições complexas, mas antes pistas e referências para exploração futura.

**PALAVRAS-CHAVE:** Bibliotecas Digitais

## INTRODUÇÃO

Este artigo fala de “bibliotecas digitais” de uma forma intencionalmente simples e directa. Não se pretendem transmitir pensamentos profundos nem conjecturas abstractas, mas antes apresentar, de uma forma que se pretende o mais clara possível, um ponto de vista pessoal sobre os assuntos em causa. As análises e conclusões são assim pessoais, claramente subjectivas, embora resultantes de uma década de experiência de vida dedicada ao tema, em que se destacam os quase seis anos como Director dos Serviços de Inovação e Desenvolvimento da Biblioteca Nacional. Estamos por isso perante um texto escrito na primeira pessoa, com que se pretende transmitir assumidamente uma mensagem pessoal.

Ainda que se estenda por um largo número de pontos, este artigo tem três objectivos, espalhados ao longo dos mesmos. O primeiro objectivo é o de apresentar, de forma sucinta aqueles que penso serem os principais conceitos que devem ser dominados por quem se pretenda debruçar de forma profunda sobre a biblioteca digital. Para além do propósito em si de explicar ou ajudar a clarificar esses conceitos, o verdadeiro

propósito deste primeiro objectivo será o de dar suporte ao segundo, distribuído por uma sequência de pontos.

O segundo objectivo é o de, através do enquadramento conceptual já referido, fornecer uma percepção do estado da arte e contar de uma forma simples uma história da biblioteca digital, com ênfase para o contexto nacional. Com isto pretende-se contribuir assim para a história das bibliotecas digitais em Portugal. Finalmente, o terceiro objectivo é o de apresentar propostas à comunidade para consideração do futuro do tema em Portugal. Este objectivo, claramente pretensioso, é quase todo concentrado no ponto final.

Antes de terminarmos este ponto, uma chamada de atenção para o triângulo bibliotecas / arquivos / museus. Desde cedo que se tem questionado o verdadeiro significado e relevância do termo “bibliotecas digitais” quando visto para além do simples contexto das bibliotecas, abrangendo igualmente os contextos dos arquivos e dos museus. Confesso que na minha perspectiva, tanto funcional como tecnológica, tenho grandes dificuldades em aceitar que quaisquer especificidades de qualquer um destes contextos possam ser consideradas suficientemente relevantes para merecer alguma compartimentação. É verdade que podemos sempre apontar alguns processos de negócio e requisitos funcionais ou não funcionais que se mostrarão mais relevantes nuns contextos que noutros, mas em termos gerais confesso que até hoje ainda não deparei com nenhum caso relevante que merecesse o esforço da compartimentação. Fica pois assim aqui a minha primeira conclusão a apresentar à comunidade!

## BREVE RESENHA HISTÓRICA (SUBJECTIVA)

A biblioteca digital deve ser vista, antes de tudo, como uma realização tecnológica. É verdade que o conceito facilmente se enquadra num imaginário acessível à generalidade das pessoas, mas no entanto o que verdadeiramente se tem feito e o que existe acaba por ser mais como sequência de acasos e reutilização de resultados externos do que o resultante de qualquer visão avisada, planeamento competente ou execução determinística.

O Memex [3] é frequentemente referido como a primeira visão de biblioteca digital, mas a verdade é que, ao que julgamos saber, pela cabeça de Vannevar Bush nunca terá passado a ideia da Internet nem da actualmente mais

genérica e poderosa das bibliotecas (os puristas que me perdoem a leviandade), a World Wide Web. Esta, por sua vez, quando surgiu como um conceito assente apenas em hiperligações unidireccionais [4], vinha violar algumas das regras fundamentais daquilo que se ensinava na altura nas universidades sobre o que deveria ser um típico sistema hipermédia. Finalmente, quando Tim Berners-Lee, o criador da World Wide Web, tentou definir o próximo passo lançando pela primeira vez a ideia da Web Semântica<sup>1</sup>, não era certamente no Google que ele estava a pensar, talvez a realização consequente mais próxima existente hoje em dia.

O caminho, na área de investigação e desenvolvimento para a realidade actual das bibliotecas digitais começou por ser equacionado na prática no início dos anos noventa do século passado, como resultado entre outros de encontros nos Estados Unidos apoiados pela NSF<sup>2</sup> [1] e de projectos promovidos pela Comissão Europeia, como o programa Telematics for Libraries [2].

A estes primeiros passos seguiu-se, nos Estados Unidos, a Iniciativa para as Bibliotecas Digitais (“Digital Libraries Initiative”), em duas fases que decorreram de 1994 a 2003<sup>3</sup>. Nas iniciativas da Comissão Europeia o tema não teve na mesma altura um contexto específico, mas um grande número de projectos relacionados têm tido acolhimento regularmente desde o 4º Programa Quadro (“4th Framework Programme”), no contexto especialmente do programa IST (“Information Society Technologies”) <sup>4</sup>. Ao mesmo tempo assistimos a alguns programas nacionais de relevo, com destaque especial por exemplo para o programa eLib no Reino Unido<sup>5</sup>.

Ao mesmo tempo que os actores públicos se movimentavam desta forma, os privados tomavam redesenhavam igualmente os seus negócios. Um exemplo notável foi dado pelos editores de conteúdos técnicos e científicos. A revolução nesta área verificada na última dezena de anos foi notável, com praticamente todos os editores de relevo a nível internacional (Elsevier, Springer, Kluwer, etc.) a desenvolverem as suas próprias bibliotecas digitais e a tentar convencer os seus clientes a abandonar o papel. Para tal alteraram mesmo os seus modelos de negócio, tornando-se comum oferecer não só o acesso aos números mais recentes das revistas, mas igualmente a todo o arquivo disponível. É assim que se pode dizer que desta forma já não há hoje em dia nenhum investigador, académico ou estudante de engenharia informática, telecomunicações ou electrotecnia que possa viver sem acesso às bibliotecas digitais da ACM<sup>6</sup> ou do IEEE<sup>7</sup>. Outro exemplo de

resultados notáveis é o serviço JSTOR – The Scholarly Journal Archive<sup>8</sup>.

Naquilo que podemos considerar como uma terceira geração deste movimento, assistimos desde 2000 ao desenvolvimento nos Estados Unidos da NSDL – National Science Digital Library<sup>9</sup>, do Programa Nacional para a Infra-estrutura de Informação e Preservação Digital (NDIIP - The National Digital Information Infrastructure and Preservation Program)<sup>10</sup> e da iniciativa dos Arquivos de Registos Electrónicos (ERA - The Electronic Records Archives)<sup>11</sup>.

Ao mesmo tempo na Europa um grupo de bibliotecas nacionais, em que se contou a Biblioteca Nacional de Portugal, conseguiu o feito concertado extraordinário (especialmente do ponto de vista político) de se associarem para o desenvolvimento do serviço da Biblioteca Europeia (TEL – The European Library)<sup>12</sup>.

Apesar da iniciativa TEL ter contado com o apoio da Comissão Europeia, que entretanto vinha financiando também quase há uma década a Rede de Excelência para as Bibliotecas Digitais DELOS<sup>13</sup>, o tema não deixou de atravessar um período de alguma hesitação e fracos resultados. Esse tempo perdido tenta-se recuperar agora, desde o lançamento em 2005 da discussão em torno da Biblioteca Digital Europeia<sup>14</sup>. Embora tenha resultado de uma reacção algo histórica às iniciativas do Google, o facto é que tal teve pelo menos o resultado de por o tema em cima da mesa na Europa, colocando-o em inesperado destaque do novo 7º Programa Quadro<sup>15</sup>, assim como no programa independente eContentplus<sup>16</sup>, isto em complemento de várias iniciativas nacionais que entretanto têm emergido localmente.

Entretanto em Portugal o tema começou por merecer atempadamente em 1997 uma entrada própria no Livro Verde para a Sociedade de Informação em Portugal, no seu ponto “3.2. Desenvolvimento de Bibliotecas Digitais” [5]. Infelizmente, tal pouco mais foi além disso mesmo, sem consequências directas relevantes. É verdade que nos últimos anos vimos surgir com interesse os resultados de alguns projectos dignos de nota, como a iniciativa da BND - Biblioteca Nacional Digital<sup>17</sup>, promovida pela Biblioteca Nacional de Portugal, e o projecto TT Online<sup>18</sup>, promovido pelo Instituto dos Arquivos Nacionais / Torre do Tombo.

Estes projectos, assim como outros de menor dimensão, têm recebido o apoio do POC – Programa Operacional

<sup>1</sup> [http://ironick.typepad.com/ironick/2006/09/history\\_of\\_the\\_.html](http://ironick.typepad.com/ironick/2006/09/history_of_the_.html)

<sup>2</sup> A NSF – National Science Foundation, é o principal organismo de financiamento à investigação científica nos Estados Unidos, desempenhando o mesmo papel que a FCT – Fundação para a Ciência e Tecnologia desempenha em Portugal.

<sup>3</sup> O arquivo desta iniciativa, que entre outros resultados deu origem ao que é hoje o serviço Google, pode ser encontrado em <http://www.dli2.nsf.gov>.

<sup>4</sup> <http://cordis.europa.eu/ist>

<sup>5</sup> <http://www.ukoln.ac.uk/services/elib>

<sup>6</sup> <http://portal.acm.org>

<sup>7</sup> <http://ieeexplore.ieee.org>

<sup>8</sup> <http://www.jstor.org>

<sup>9</sup> <http://nsdl.org/>

<sup>10</sup> <http://www.digitalpreservation.gov>

<sup>11</sup> <http://www.archives.gov/era>

<sup>12</sup> <http://www.theeuropeanlibrary.org>

<sup>13</sup> <http://www.delos.info>

<sup>14</sup> <http://europa.eu.int/idabc/en/document/4239/330;>

<http://dossiers.publico.pt/shownews.asp?id=1222269&idCanal=10>

<sup>15</sup> [http://cordis.europa.eu/fp7/ict/programme/challenge4\\_en.html](http://cordis.europa.eu/fp7/ict/programme/challenge4_en.html)

<sup>16</sup> [http://europa.eu.int/information\\_society/activities/econtentplus](http://europa.eu.int/information_society/activities/econtentplus)

<sup>17</sup> <http://bnd.bn.pt>

<sup>18</sup> <http://ttonline.iantt.pt>

da Cultura<sup>19</sup> e do POS\_Conhecimento – Programa Operacional da Sociedade do Conhecimento<sup>20</sup>. Mas todo este contexto se tem desenvolvido de modo não articulado, permitindo marcações de posição individuais, mas praticamente quase só simbólicas, desligadas de qualquer dinâmica de sinergias ou de sustentação.

O papel do poder político nesta área em Portugal tem sido assim tradicionalmente ausente. Tem-se verificado uma postura consentida perante as iniciativas dos actores no terreno (que as tentam desenvolver “como podem”), mas nada disso tem sido enquadrado em qualquer política científica, cultural ou de cidadania nacional. Como excepção deve ser registada no entanto, numa nota claramente à parte, a iniciativa B-ON<sup>21</sup>, a qual permitiu à comunidade académica nacional passar a aceder de forma cómoda aos principais títulos de referência internacionais. Em complemento a estas referências podíamos apontar mais casos interessantes, não só exclusivamente da área das bibliotecas mas também dos arquivos e museus. No entanto, regra geral esses resultados têm sido localizados, não susceptíveis de se replicarem facilmente ou de se constituírem como paradigmas genéricos (especialmente nos casos de projectos nacionais<sup>22</sup>).

De seguida iremos listar e definir brevemente os principais conceitos chave que, na minha perspectiva devem ser levados em consideração e dominados minimamente por quem pretende desenvolver um conhecimento relevante nesta área.

Alguns dos conceitos apresentados são “casos de uso”, uma designação comum em engenharia, na modelação de processos ou de sistemas, para referir e descrever processos de negócio ou actividades de onde devem resultar requisitos para o desenvolvimento dos sistemas.

## **SOBRE DIGITALIZAÇÃO E EDIÇÃO DIGITAL**

Os casos da digitalização e da edição digital têm como objectivo a criação de objectos digitais.

A **digitalização** pode-se definir genericamente como qualquer processo de transposição de obras impressas que pode resultar em imagem, texto (através de transcrição), som (livros falados), etc. Já a **edição digital** deve resultar na produção de obras originais em formato digital, incorporando ou não conteúdos digitalizados, mas com valor acrescentado e identidade bibliográfica própria. Estas obras podem consistir na publicação digital de uma obra já existente noutro meio (como por exemplo o texto, em formato digital, de um livro de Almeida Garrett), ou uma publicação digital original, feito ou não com conteúdos digitalizados, como são por

---

<sup>19</sup> <http://poc.min-cultura.pt>

<sup>20</sup> <http://www.posc.mctes.pt>

<sup>21</sup> <http://www.b-on.pt/>

<sup>22</sup> Por exemplo, uma iniciativa nacional que aparenta ser interessante mas que se apresenta enigmática pela falta de divulgação e dificuldades postas ao acesso é a Nesos – Base de Dados das Ilhas Atlânticas (<http://www.nesos.net/>).

exemplo os dossiers temáticos do Público<sup>23</sup> ou da Fundação Mário Soares<sup>24</sup>.

A edição digital seria, à partida, um caso estranho aos contextos tradicionais das bibliotecas, arquivos ou museus, pois pertencem ao mundo externo dos criadores. No entanto pelo menos nestas fases iniciais em que nos encontramos ele torna-se pertinente, especialmente associado ao caso da digitalização, em que as organizações se têm mobilizado para converter os seus espólios físicos para digital. Neste contexto as bibliotecas, arquivos e museus acabam por partilhar problemas e dificuldades com os criadores externos, motivando à partilha de conhecimento e à definição conjunta de novas normas e de boas práticas, e mesmo de novos modelos de negócio.

Alguns projectos relevantes têm sido levados a cabo em Portugal nesta área, financiados pelo POC ou POSI (como nos casos da BND, TT Online, Arquivo Mário Soares, etc.) ou com recursos próprios (como é o caso extremamente interessante mas infelizmente pouco divulgado da iniciativa pioneira da Assembleia da República de digitalização dos Debates Parlamentares<sup>25</sup>, e ainda o caso mais recente do Fundo Antigo da Faculdade de Ciências da Universidade do Porto<sup>26</sup>).

Apesar destes e outros casos pontuais, o cenário de sinergia nacional é pouco relevante. Os projectos têm-se desenvolvido em grande parte de forma isolada, sem tradição de comunicação entre si nem tentativas de criação de condições para tal. Não se tem procurado assegurar a partilha e transferência de conhecimento de umas iniciativas para as outras (tanto por quem faz como por quem pretende fazer), pelo que assistimos à emergência de resultados espúrios, sem coordenação nem normalização entre si.

No caso dos projectos de digitalização financiados pelo POC, este problema foi apontado pela BN na altura da segunda leva de contratações. Tendo sido ainda possível garantir a partir daí alguma coordenação com a BND, incluindo a sensibilização das entidades executantes para depositarem os seus resultados na BN. No entanto era já tarde demais para garantir uma actividade adequada, especialmente para formação, pelo que só já foi possível promover alguma coordenação posterior e tentar remediar situações herdadas.

Visto o POS\_Conhecimento também ter apoiado projectos de criação de conteúdos, a mesma aproximação foi feita a este programa (na altura ainda designado POSI). A resposta recebida, dada pessoalmente, foi algo caricata e bastante reveladora da mais pura perspectiva burocrática nacional. Entendeu o POSI que a BN lhe deveria remeter as suas observações e propostas através da tutela, senão do Ministério da Cultura, então através da Secretaria-geral do mesmo, como entidade de ligação ao POSI. Se um problema se

---

<sup>23</sup> <http://dossiers.publico.pt/>

<sup>24</sup> [http://www.fmsoares.pt/arquivo\\_biblioteca/004\\_dossiers\\_apresenta.asp](http://www.fmsoares.pt/arquivo_biblioteca/004_dossiers_apresenta.asp)

<sup>25</sup> <http://debates.parlamento.pt/>

<sup>26</sup> <http://www.fc.up.pt/fa/>

pode resolver em dois passos, ou mesmo em três, porque fazê-lo apenas num?

## SOBRE OBJECTOS DIGITAIS

Foi dito atrás que os casos da digitalização e da edição digital têm como objectivo a criação de objectos digitais. O que são então objectos digitais? Pode parecer estranho afirmar isto, mas este conceito fundamental nem sempre tem sido claramente entendido.

De forma simples, um **objecto digital** pode ser definido neste contexto como qualquer recurso de informação que possa ser descrito e referenciado nas perspectivas bibliográficas, arquivísticas ou museológicas.

É fácil e imediato para todos reconhecer como exemplo imediato de um objecto digital algo que esteja gravado num CD-ROM ou DVD, mas com a desmaterialização dos processos e a afirmação da Internet como meio preponderante, há que aprofundar o conceito.

Um conceito igualmente imediato mas já liberto de constrangimentos físicos, será o E-book. Este termo deriva da designação inglesa “Electronic Book”, ou livro electrónico, referindo-se a um livro em formato digital que pode ser lido em computadores, PDAs ou mesmo telemóveis. Este objecto tenta reproduzir, em todas as suas características básicas, as características de um livro convencional, embora normalmente contenha também características avançadas como hiperligações, anotações electrónicas, etc. Um outro conceito de objecto digital facilmente identificável e também já vulgar hoje em dia é o dos ficheiros de áudio, de música, em formato MP3 ou outros, que descarregamos de serviços especializados, fazemos circular pela Internet, armazenamos nos nossos computadores ou leitores pessoais, etc.

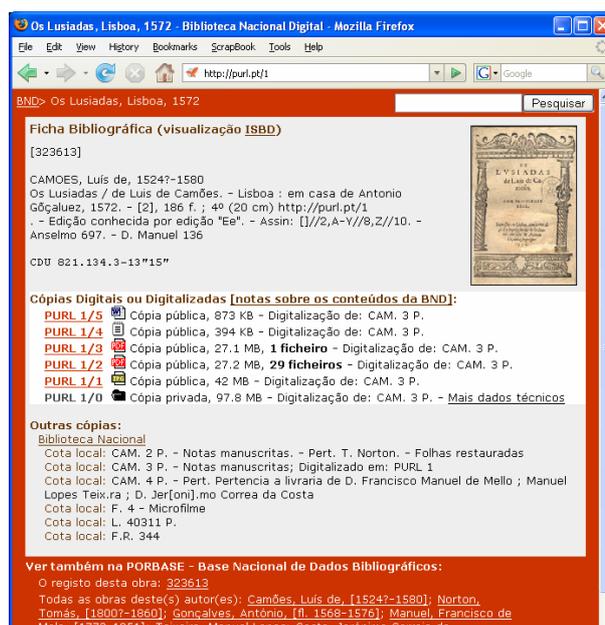
O conceito de E-book é facilmente entendido porque é geralmente apresentado “encadernado” tecnicamente num único ficheiro (em formato MS-Word, PDF, ou outro equivalente). Por essa razão é igualmente frequente ver-se exigir como resultado de projectos de digitalização a entrega dos resultados finais em PDF. Este é um requisito cómodo, mas que nem sempre se adequa a todos os cenários de utilização. Como consequência corremos o risco de acabar com ficheiros de tamanhos monstruosos, possíveis de serem transferidos em redes locais, mas quase impossível de transmitir pela Internet.

Vamos parar assim ao conceito de objecto digital complexo. Neste caso o objecto poderá ser uma estrutura de vários ficheiros, explicitamente estruturados como um “sítio web”, ou implicitamente descrito através de metadados estruturais. O primeiro caso tem a vantagem de permitir ainda manipular e referenciar o objecto de forma fácil e independente da tecnologia (todo o objecto pode estar organizado dentro de uma mesma pasta no sistema de ficheiros, e pode-se ter um URL para a sua página de entrada). Já o segundo caso se mostra mais complexo, pois exige que nalgum momento os

metadados estruturais sejam manipulados por um sistema de informação para dar acesso aos conteúdos de informação, recorrendo-se eventualmente a uma base de dados para tal.

A opção por sistemas de gestão de conteúdos suportados em bases de dados, para armazenamento de imagens, vídeo, som ou texto, faz todo o sentido em contextos de grande dinamismo ou ritmo de produção, como por exemplo num jornal diário. Já no contexto de uma biblioteca digital, onde a estabilidade dos objectos é um dos objectivos, tal opção pode ter consequências desastrosas. É que numa solução dessas o que se referencia e manipula não são os objectos no seu todo, mas as suas partes, correndo-se assim o risco de facilmente se perder a noção do conceito de objecto. Talvez este alerta seja mais claro para os nossos colegas arquivistas, se disser que é como ter um arquivo permanente gerido por regras e processos de um arquivo corrente. Ainda neste contexto toda a biblioteca digital acaba por ser assim não uma colecção de objectos mas um gigantesco objecto em si. Os acessos e uso das obras ficam dessa forma completamente dependentes do sistema de informação, criando dificuldades à sua cópia, identificação e preservação.

No caso da BND este problema foi abordado de modo a se tentar acolher da melhor forma todos os requisitos. Assim, todos os objectos são organizados **objectos de preservação**, resultando em cada caso numa pasta contendo todos os ficheiros com os conteúdos de informação e ainda os respectivos metadados estruturais. Assim, é possível em qualquer momento manipular esses conteúdos, extraíndo o todo ou partes. Já para acesso e referência foram criadas outras cópias mais adequadas a partir dos objectos de preservação, na forma de estruturas HTML, ficheiros PDF, ou o que se mostrar mais conveniente para cada caso. A Figura 1 mostra um exemplo de um objecto da BND (“Os Lusíadas”) nessas condições, oferecendo-se várias opções de **objectos de acesso**.



The screenshot shows a web browser window with the address bar displaying 'http://purl.pt/1'. The page content includes:

- Ficha Bibliográfica (visualização ISBD)** [323613]
- CAMOES, Luis de, 1524?-1580**
- Os Lusíadas / de Luis de Camões. - Lisboa : em casa de Antonio Gõçgalvez, 1572. - [2]. 186 f. ; 4º (20 cm) http://purl.pt/1**
- Edição conhecida por edição "Ee". - Assin: [1]/2,A-Y//8,2//10. - Anselmo 697. - D. Manuel 136**
- CDU 821.134.3-13"15"**
- Cópias Digitais ou Digitalizadas [notas sobre os conteúdos da BND]:**
  - PURL 1/5** Cópia pública, 873 KB - Digitalização de: CAM. 3 P.
  - PURL 1/4** Cópia pública, 394 KB - Digitalização de: CAM. 3 P.
  - PURL 1/3** Cópia pública, 27.1 MB, **1 ficheiro** - Digitalização de: CAM. 3 P.
  - PURL 1/2** Cópia pública, 27.2 MB, **29 ficheiros** - Digitalização de: CAM. 3 P.
  - PURL 1/1** Cópia pública, 42 MB - Digitalização de: CAM. 3 P.
  - PURL 1/0** Cópia privada, 97.8 MB - Digitalização de: CAM. 3 P. - [Mais dados técnicos](#)
- Outras cópias:**
  - Biblioteca Nacional**
    - Cota local: CAM. 2 P. - Notas manuscritas. - Pert. T. Norton. - Folhas restauradas
    - Cota local: CAM. 3 P. - Notas manuscritas; Digitalizado em: PURL 1
    - Cota local: CAM. 4 P. - Pert. Pertencia a livraria de D. Francisco Manuel de Mello ; Manuel Lopes Teixeira ; D. Jer[on]mo Correa da Costa
    - Cota local: F. 4 - Microfilme
    - Cota local: L. 40311 P.
    - Cota local: F.R. 344
- Ver também na PORBASE - Base Nacional de Dados Bibliográficos:**
  - O registo desta obra: [323613](#)
  - Todas as obras deste(s) autor(es): [Camões, Luis de, \[1524?-1580\]](#); [Norton, Tomás, \[1800?-1860\]](#); [Gõçgalves, António, \[fl. 1568-1576\]](#); [Manuel, Francisco de Mello, \[1773-1851\]](#); [Teixeira, Manuel Lopes, Costa, Jeronimo Correa da](#)

Figura 1: Exemplo de uma obra digitalizada com várias cópias de acesso.

## **SOBRE DEPÓSITO, AQUISIÇÃO E LICENCIAMENTO**

O caso da **aquisição** compreende as acções legais, comerciais e técnicas necessárias para que uma entidade possa entrar na posse física de uma publicação. As bibliotecas públicas já estão habituadas há muito a construir colecção multimédia, em cassetes VHS, CD-ROM ou DVD, mas esse cenário está evoluindo de forma bastantes significativa.

Já se tornou habitual para muitas bibliotecas universitárias oferecer aos seus utilizadores colecções de E-books, seja por aquisição directa dos mesmos (que podem assim ser instalados num sistema de armazenamento local) ou apenas licenciando um serviço de acesso remoto ao repositório do editor ou de algum intermediário. Este cenário de licenciamento é o que foi concretizado com a B-ON.

Na BND foram desenvolvidos processos e sistemas que permitem a constituição de colecções por aquisição de obras digitais ou digitalizadas, considerando vários cenários. O serviço RECOLHA suporta processos de aquisição despoletados pela BN, aplicadas a obras ou recursos em geral na Internet. Foram assim constituídas desta forma em 2005 e 2006 interessantes colecções de recolha de sítios de partidos políticos, candidatos, primeiras páginas de câmaras municipais e de jornais, relativos aos actos eleitorais para a Assembleia da República, Autarquias Locais, e Presidência da República.

O depósito de obras numa biblioteca digital poderá efectuar-se por obrigação legal, contratual ou de modo voluntário. No caso da BND temos assim o exemplo do serviço DITED – Depósito de Teses e Dissertações Digitais<sup>27</sup>, e na Universidade do Minho o caso do REPOSITORIUM<sup>28</sup>, alinhado com os objectivos do movimento de promoção de **repositórios institucionais**<sup>29</sup>.

Um problema relevante que não deve ser descurado neste contexto é o da garantia da **autenticidade** dos objectos. O problema não é simples, mas deverá ser encarado em cada caso com os seus requisitos específicos. No caso dos depósitos no sistema DITED, por exemplo, o processo de cada tese inclui contactos com o autor, que ficam registados ao mesmo (podem ser mensagens de correio electrónico). No final, aos objectos desta forma depositados (tal como todos os objectos da BND) são adicionadas assinaturas electrónicas às suas cópias.

Este simples mecanismo não é isento de falhas, mas responde de forma eficaz ao requisito principal que é o de proteger as obras não tanto contra atentados intencionais à sua integridade, mas sim contra erros nos processos ou sistemas. Já para um caso mais exigente (como por exemplo num arquivo permanente a depositar nos Arquivos Nacionais), ainda que tecnicamente o processo usado pudesse ser semelhante, a sua

concretização deveria ser mais robusta. No entanto não é nada que não encontre já resposta hoje em dia na tecnologia corrente.

## **SOBRE DESCRIÇÃO E REGISTO**

O **registo** compreende o conjunto de tarefas relacionadas com a recolha, organização e gravação de informação sobre um objecto de informação necessária para a sua descrição, gestão, procura, acesso e preservação. Este processo compreende pois as actividades tradicionais de catalogação, podendo ainda, consoante o contexto, incluir outras actividades (tal como o registo de informação de proveniência, localização, licenciamento, autenticidade, etc.). Dito de outra forma, este caso destina-se essencialmente a criar **metadados**.

Tradicionalmente as bibliotecas têm recorrido a formatos da família MARC para este propósito, mas recentemente uma nova iniciativa tem vindo a tentar criar novas perspectivas mais adequadas para o mundo digital.

Trata-se da iniciativa RDA – Resource Description and Access<sup>30</sup> [6], que pretende conduzir ao desenvolvimento de novas regras e normas para a descrição e acesso a recursos em contexto digital. Pretende ser uma actividade baseada no trabalho anterior das AACR – Anglo-American Cataloguing Rules (em que as Regras Portuguesas de Catalogação se baseiam), e ainda que seja levada a cabo na perspectiva das bibliotecas, o seu propósito é mais alargado, devendo resultar em propostas compatíveis com propostas equivalente dos arquivos, museus e editores.

A iniciativa RDA leva em consideração o modelo já estabelecido para os FRBR – Functional Requirements for Bibliographic Records<sup>31</sup>, assim como outros relacionados para outras áreas, tal como o CIDOC Conceptual Reference Model<sup>32</sup> (museus e área cultural em geral), ONIX – ONline Information eXchange<sup>33</sup> (editores), etc.

Um aspecto especialmente importante desta iniciativa é a ênfase dado a cenários de reutilização de metadados. Efectivamente assume-se que, perante o cenário emergente de criação alargada de recursos de informação das mais variadas formas e origens, não será mais possível às bibliotecas e instituições relacionadas continuar a produzirem por si informação de descrição como até agora. Terá sido já essa constatação que levou em parte ao desenvolvimento do formato Dublin Core<sup>34</sup>, concebido inicialmente apenas como um formato para responder a requisitos de **pesquisa** (“resource discovery”) em cenários de reutilização de elementos de metadados criados originalmente noutros contextos. Posteriormente o mundo do Dublin Core terá vindo a ser

<sup>27</sup> <http://dited.bn.pt>

<sup>28</sup> <https://repositorium.sdum.uminho.pt/>

<sup>29</sup> [http://en.wikipedia.org/wiki/Institutional\\_repository](http://en.wikipedia.org/wiki/Institutional_repository)

<sup>30</sup> <http://www.collectionscanada.ca/jsc/rdapropectus.html>

<sup>31</sup> [www.ifla.org/VII/s13/frbr/frbr.pdf](http://www.ifla.org/VII/s13/frbr/frbr.pdf)

<sup>32</sup> <http://cidoc.ics.forth.gr/>

<sup>33</sup> <http://www.editeur.org/>

<sup>34</sup> <http://dublincore.org/>

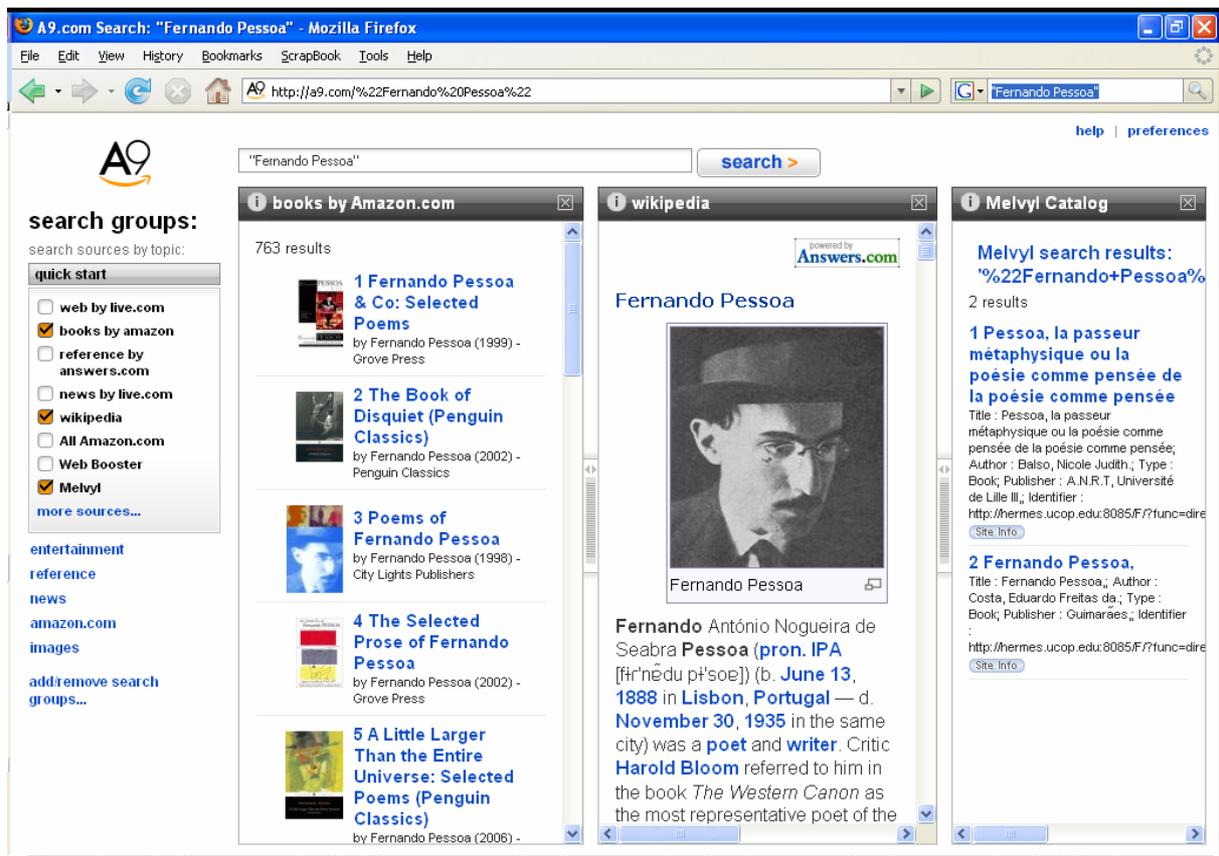


Figura 2: Interface do motor de busca A9, para um exemplo de pesquisa por “Fernando Pessoa” e com resultados para os espaços Amazon.com, Wikipedia e o catálogo das bibliotecas da Universidade da Califórnia (Melvyl).

alargado com a definição de qualificadores, perfiz, etc., mas sempre se considerou que essa não seria a melhor solução para garantir uma adequada descrição dos recursos.

Foi esse motivo que terá levado a Biblioteca do Congresso nos Estados Unidos a definir o MODS – Metadata Object Description Schema, uma versão simplificada do MARC21<sup>35</sup> para uso nos seus projectos de biblioteca digital.

Entretanto outras comunidades têm vindo a definir também os seus esquemas de **metadados descritivos** (redes de rádios, televisão, arquivos de cinema, etc.). Espera-se assim que quando essas decisões se demonstrarem venham a contribuir para cenários complexos de pesquisa de informação, coexistindo com as bibliotecas, museus e arquivos em geral.

## SOBRE PROCURA

A **procura** refere-se aos processos de descoberta dos objectos que respondam a um determinado conjunto de critérios (pesquisa bibliográfica, navegação em índices, etc.). Podem-se incluir ainda neste caso cenários de disseminação selectiva de informação. A ferramenta tradicional para este objectivo tem sido o OPAC – On-line Public Access Catalogue. Esta é no entanto uma

ferramenta já claramente desactualizada para a realidade actual.

A perspectiva principal do OPAC recorre a uma visão de uso da tecnologia assente em termos gerais no conceito de formulário de pesquisa (a célebre opção de “pesquisa avançada” que tanto nos habituámos a ver). Este modelo foi adequado aos primeiros tempos dos sistemas de informação e bases de dados, em que as estruturas de dados eram forçosamente fixas e o utilizador se via na necessidade de expressar de forma complexa a sua fórmula de pesquisa.

Por outras palavras, sendo o OPAC um conceito substituído do catálogo manual, era natural que fosse concebido segundo as mesmas perspectivas. Esses conceitos tornam-se no entanto hoje em dia um estorvo, em que podemos perfeitamente delegar na máquina o ónus do esforço. Isto é, os novos modelos de procura pretendem tirar partido das grandes capacidades e velocidade dos computadores actuais para, pedindo ao utilizador um conjunto mínimo de elementos, oferecer-lhe de volta não apenas um alista de resultados, mas todo um conjunto de espaços variados e alternativos.

Um interessante exemplo deste paradigma é o sistema de pesquisa A9<sup>36</sup>, ilustrado na Figura 2. Um conceito semelhante está a ser desenvolvido no sistema de pesquisa MITRA, que neste momento suporta o serviço

<sup>35</sup> <http://www.loc.gov/standards/mods/>

<sup>36</sup> <http://a9.com>

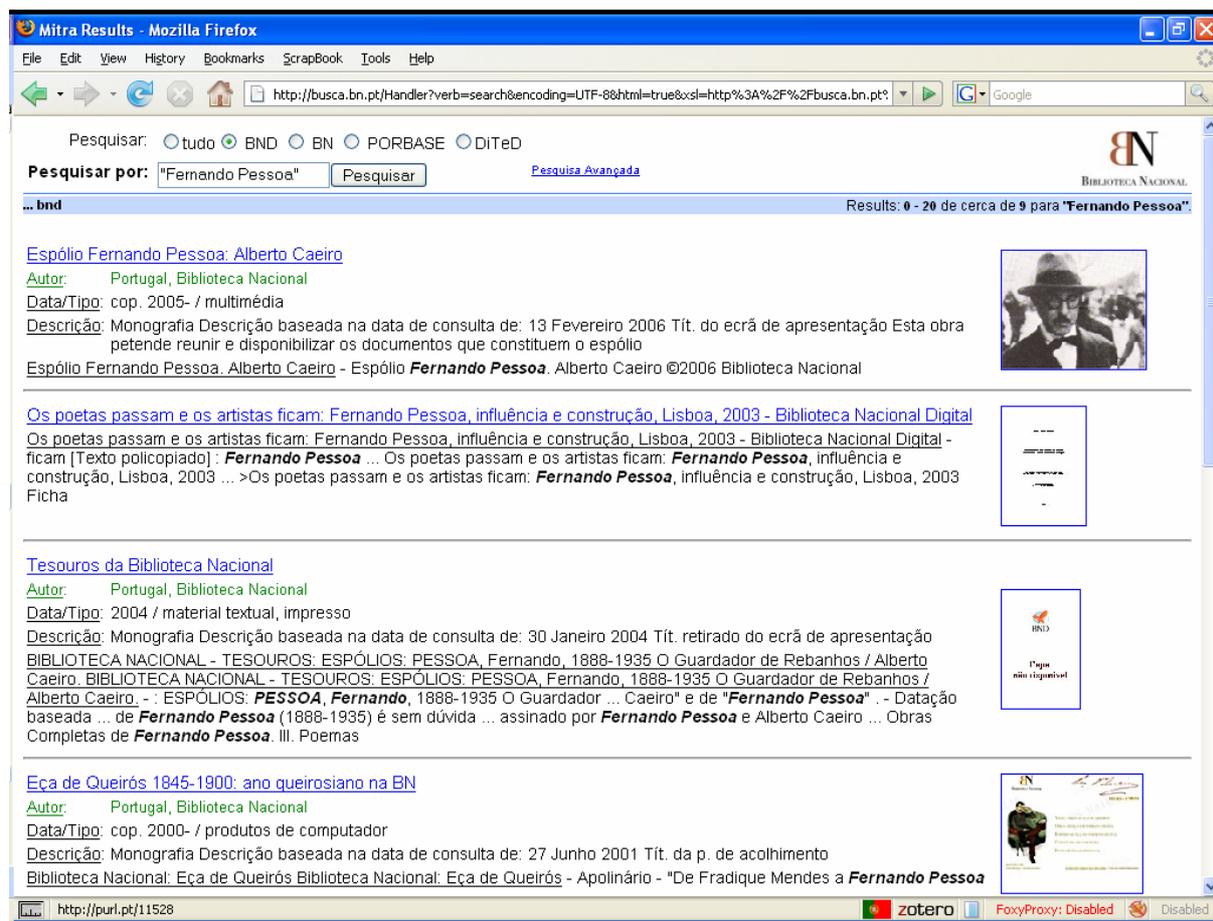


Figura 3: Interface do motor de busca da BND, para um exemplo de pesquisa por “Fernando Pessoa”.

BUSCA<sup>37</sup> da BN (protótipo mostrado na Figura 3, com possibilidade de pesquisa na BND, DITED e sítios WEB da BN e PORBASE). Um conceito semelhante está sendo concretizado no serviço Google, com interligação entre o serviço original e os novos serviços Google Scholar e Google Book Search.

## SOBRE ACESSO

O **acesso** é o processo que garante a um determinado utilizador usufruir de um objecto, na forma de um item ou conjunto de itens que definem o conteúdo intelectual desse recurso.

Uma contribuição relevante para este caso é a infraestrutura **OpenURL** – The OpenURL Framework for Context-Sensitive Services<sup>38</sup>. O objectivo desta solução é, para um dado utilizador de um sistema que terminou um processo de pesquisa e identificou um objecto a que pretende aceder, permitir ao mesmo indagar agora onde poderão eventualmente existir cópias desse objecto para acesso. Segundo a identificação do utilizador e as alternativas de licenciamento a que o mesmo possa ter direito, o sistema poderá assim responder, para o mesmo objecto, com destinos e condições de acessos diversas.

Outra realidade emergente recente é o movimento

<sup>37</sup> <http://busca.bn.pt>

<sup>38</sup> [http://www.niso.org/committees/committee\\_ax.html](http://www.niso.org/committees/committee_ax.html)

Acesso Livre (**Open Access**<sup>39</sup>). O propósito deste movimento é o de promover a emergência de conteúdos de ciência e tecnologia de qualidade em acesso livre, tais como teses, relatórios e artigos científicos. Um exemplo de um recurso valioso associado a este contexto é a Directory of Open Access Journals<sup>40</sup>.

Movimentos desta natureza levantam o problema da declaração de opções de termos e condições para acesso e uso dos objectos. É assim neste contexto que vimos emergir recentemente a iniciativa **Creative Commons**<sup>41</sup>, já com uma instância para Portugal<sup>42</sup>.

## SOBRE ARMAZENAMENTO

O armazenamento compreende todos os processos e técnicas destinadas a colocar uma publicação num local onde a mesma possa ser acedida (segundo as suas regras específicas ou outras gerais).

Os custos da tecnologia de armazenamento têm descido drasticamente nos últimos tempos, permitindo perspectivar um futuro risonho para projectos de digitalização de alta qualidade. Ainda é necessário manter cuidados especiais com a qualidade dos discos

<sup>39</sup> [http://en.wikipedia.org/wiki/Open\\_access](http://en.wikipedia.org/wiki/Open_access)

<sup>40</sup> <http://www.doaj.org/>

<sup>41</sup> <http://creativecommons.org/>

<sup>42</sup> <http://creativecommons.org/worldwide/pt/>

vulgares, mas os custos baixos dos mesmos permitem desenhar com algum cuidado estratégias de replicação eficazes.

## SOBRE PRESERVAÇÃO

A **preservação digital** compreende todos os processos destinados a garantir a boa forma física, lógica e intelectual da publicação, tendo como objectivo último garantir o seu acesso em qualquer momento futuro.

O projecto NEDLIB<sup>43</sup>, em que a BN e o INESC foram parceiros, foi pioneiro na abordagem a este problema, com um encontro de grande relevância realizado em Portugal, em Tomar, em Junho de 1998. Foi um projecto onde muito se trabalhou, muito se aprendeu, mas pouco se produziu. No entanto, algo de muito relevante se aprendeu na altura, e que entretanto tem vindo a ser confirmado por outras iniciativas: é que o combate ao problema da preservação se encontra mais na definição das políticas e dos processos do que na tecnologia.

Uma proposta que surgiu ao mesmo tempo que se desenvolvia o projecto NEDLIB foi a do modelo de referência para preservação OAIS – Reference Model for an Open Archival Information System<sup>44</sup>. Surgindo num momento de desespero da comunidade pela falta de resultados e consensos das várias iniciativas em curso, foi uma proposta imediatamente recebida de braços abertos como salvadora. Este foi um momento feliz para os decisores associados ao problema, pois permitiu-lhes uma sensação de missão cumprida, mas é preciso no entanto não nos deixarmos enganar pela sua excessiva simplicidade e constatação do óbvio. Tecnicamente é mesmo uma proposta muito pouco relevante.

Foi por se reconhecerem todos estes aspectos delicados que no contexto da BND sempre se afirmou não se estar a levar em linha de conta o problema da preservação a longo prazo. Todo o trabalho desenvolvido teve sempre como preocupação máxima o caso do armazenamento. Ainda assim, garantiu-se a estruturação correcta dos objectos para uma manipulação flexível em qualquer cenário actual ou futuro, pelo que em qualquer momento se poderão definir e aplicar eficazmente políticas de preservação adequadas. Mas não seria correcto afirmar neste momento que tal tenha sido definido ou esteja a ser feito.

Entretanto os Arquivos Nacionais iniciaram o projecto RODA<sup>45</sup>, uma iniciativa abordando entre outros o problema da preservação. É claramente uma iniciativa honesta, mas algo confusa, em que aparentemente se misturam requisitos funcionais com requisitos não funcionais de acordo com critérios de prioridade pouco claros. Claramente, há um grande risco em por um lado se abordarem problema complexos com aproximações excessivamente simplistas, ao mesmo tempo que se tentam abordar um número excessivamente alargado de

tópicos para se conseguir garantir um conjunto coerente de resultados. Sinto-me perfeitamente à vontade para proferir estas afirmações, pois foram precisamente estes os erros mais relevantes cometidos no desenho e desenvolvimento da BND enquanto sob minha responsabilidade (objectivos nem sempre claros e excessivamente ambiciosos para os recursos disponíveis).

## SOBRE INTEROPERABILIDADE

Ourto conceito fundamental associado ao paradigma da biblioteca digital é o da interoperabilidade. Escrito desta forma, o termo é um neologismo do Inglês “interoperability”, que pretende designar a capacidade de dois ou mais sistemas poderem comunicar entre si.

Esta **interoperabilidade espacial**<sup>46</sup> traduz-se essencialmente na capacidade de os dois sistemas trocarem informação segundo esquemas de metadados e protocolos entendidos por todos os envolvidos, permitindo assim o desenvolvimento de serviços unificados mas utilizando recursos de variadas fontes.

Um exemplo consequente destes processos é o portal TEL, assente em normas de interoperabilidade como o formato de metadados Dublin Core (tendo sido definido um perfil próprio para o serviço TEL) e os protocolos OAI-PMH<sup>47</sup>, Z39.50<sup>48</sup> e SRU/SRW<sup>49</sup>.

## SOBRE METADADOS

**Metadados** são todas as estruturas de informação que se utilizem numa biblioteca digital para registar qualquer tipo de informação relevante sobre os objectos. O termo metadados pode ser usado de forma variada em variados contextos, pelo que é necessário tomar atenções especiais quando se pretende fazer-lhes referência no contexto da biblioteca digital<sup>50</sup>.

Temos assim esquemas de metadados descritivos (UNIMARC<sup>51</sup>, MARC21<sup>52</sup>, ONIX, CIDOC, EAD<sup>53</sup>, etc.), metadados para pesquisa (Dublin Core, TEL-profile, etc.), estruturais (METS<sup>54</sup>, MPEG-21<sup>55</sup>, etc.), preservação (PREMIS<sup>56</sup>, etc.), etc.

<sup>43</sup> <http://nedlib.kb.nl/>

<sup>44</sup> [http://ssdoo.gsfc.nasa.gov/nost/isoas/ref\\_model.html](http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html)

<sup>45</sup> <http://roda.ianntt.pt>

<sup>46</sup> Em oposição ao termo “interoperabilidade no tempo” por vezes usado para designar também o caso da preservação.

<sup>47</sup> <http://www.openarchives.org/>

<sup>48</sup> <http://www.loc.gov/z3950/agency/>

<sup>49</sup> <http://www.loc.gov/standards/sru/>

<sup>50</sup> <http://metadados.bn.pt>

<sup>51</sup> <http://www.unimarc.info>

<sup>52</sup> <http://www.loc.gov/marc>

<sup>53</sup> <http://www.loc.gov/ead> (sob certa perspectiva, este caso poderá considerar-se também um esquema estrutural).

<sup>54</sup> <http://www.loc.gov/standards/mets>

<sup>55</sup> <http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm>

<sup>56</sup> <http://www.oclc.org/research/projects/pmwg/>

## CONCLUSÃO

Termino com uma lista de considerações e propostas à comunidade. Em todas elas me afirmo ao mesmo tempo como autor e alvo. É para mim motivo de grande prazer poder-me apresentar como membro da BAD, por isso evoco o direito de falar abertamente:

- É importante promover uma **iniciativa nacional concertada para a preservação digital**. Este é um problema muito complexo, que requer a afirmação de uma entidade de referência a nível nacional, seja física ou virtual – não se crê que existam especificidades entre as bibliotecas, arquivos e museus, pelo que todos estes âmbitos deveriam ser enquadrados no mesmo contexto. Esta iniciativa deveria ser promovida por exemplo em parceria entre o Ministério da Cultura e o Ministério da Ciência e Ensino Superior.
- É importante promover uma **iniciativa nacional para os repositórios institucionais**, idealmente a enquadrar com a B-ON e com um apoio claro do Ministério da Ciência e Ensino Superior.
- É importante promover uma **iniciativa nacional para a digitalização**, em parceria por exemplo entre o Ministério da Cultura, o Ministério da Ciência e Ensino Superior, e o Ministério da Educação, visando o desenvolvimento de uma biblioteca digital significativa representativa da nossa história, cultura e ciência. Uma actividade de interesse especial neste contexto seria um programa nacional de digitalização de periódicos.
- É importante promover um **serviço nacional para recolha de metadados** e com base nele um **serviço nacional de pesquisa e descoberta de recursos culturais, históricos e científicos**, a levar a cabo por exemplo por parceria entre o Ministério da Cultura, o Ministério da Ciência e Ensino Superior, e o Ministério da Educação. Por outras palavras, deveriam replicar-se em Portugal os objectivos da European Digital Library, permitindo assim ao país não só apresentar um serviço nacional relevante, mas também apresentar-se com uma frente unida perante os parceiros internacionais

Em complemento, realço ainda os seguintes alertas:

- Voltando ao triângulo bibliotecas / arquivos / museus, e agora não só numa perspectiva de massa crítica, mas também de sinergia, é urgente promover uma colaboração funcional mais estreita, especialmente em relação ao desenvolvimento de serviços de descoberta de recursos e de portais. Esta é uma tendência internacionalmente cada vez mais comum<sup>57</sup>, em que os resultados só podem ser positivos. Foi pena a oportunidade perdida da iniciativa prevista mas não executada do Portal da Cultura...

<sup>57</sup>Encontramos organismos MLA – Museums, Libraries and Archives em várias regiões ou mesmo a nível nacional no Reino Unido, Estados Unidos, Canadá, Noruega, etc.

- Cada um por si é sempre má ideia! É realmente triste ver como num país tão pequeno temos tanta dificuldade em nos encontrarmos, falarmos, trocarmos experiências e discutirmos objectivos comuns. É confrangedor o estado quase permanente de desconfiança mútua, quer entre indivíduos quer entre organizações. O resultado deste estado de “costas viradas” é o repetir dos erros e a ineficiência na execução. A quase totalidade das iniciativas na área das bibliotecas digitais em Portugal tem sido financiada com dinheiros públicos, pelo que penso que devíamos todos respeitar isso melhor e reconsiderar as nossas posturas tradicionais.
- Devíamos evitar repetir os erros do POC e do POSI / POS\_Conhecimento. Em programas com tamanha importância e capacidade de investimento devia ser obrigatório preparar previamente o terreno para os agentes beneficiários. No caso concreto dos projectos nas áreas de digitalização e desenvolvimento de conteúdos, é lamentável não se terem planeado acções prévias de formação e normalização. Como consequência, é de recear que grande parte do investimento feito se venha a perder (se não se perdeu já).
- É urgente cooperar mais e promover com maior ênfase os conceitos de comunidade, parceria e massa crítica. A dimensão típica de um projecto em Portugal não permite geralmente tirar vantagem dos factores de escala, pelo que deviam ser equacionadas formas de trabalhar alternativas mais eficazes. A iniciativa da B-ON é um exemplo interessante do que se pode conseguir quando promove massa crítica! A PORBASE é outro exemplo que devia ser acarinhado (neste caso, mais do que uma realidade consequente, estamos mais perante um espaço subaproveitado).

## NOTAS

- [1] Fox, E. – Digital Libraries Source Book. [Em linha] 1993. [Consult. 4 Fev. 2007]. Disponível em WWW: <<http://fox.cs.vt.edu/DLSB.html>>
- [2] Comissão Europeia - Telematics for Libraries [Em linha] 1998. [Consult. 4 Fev. 2007]. Disponível em WWW: <<http://cordis.europa.eu/libraries/>>
- [3] Bush, V. – As We May Think. The Atlantic Monthly [Em linha] Julho 176:1 (1965), 101-108. [Consult. 4 Fev. 2007]. Disponível em WWW: <<http://www.theatlantic.com/doc/194507/bush>>
- [4] Berners-Lee, T. – Weaving the Web: The Past, Present and Future of the World Wide Web by its Inventor. Texere, 2000. ISBN 1-58799-018-0
- [5] Missão para a Sociedade da Informação – Livro Verde para a Sociedade da Informação em Portugal [Em linha] 1997. [Consult. 4 Fev. 2007]. Disponível em WWW: <<http://purl.pt/239>>
- [6] Coyle, K., Hillmann, D. – Resource Description and Access (RDA): Cataloguing Rules for the 20th Century. [Em linha] D-Lib Magazine, 13:1/2 January/February 2007. [Consult. 4 Fev. 2007]. Disponível em WWW: <http://www.dlib.org/dlib/january07/coyle/01coyle.html>