

UNIDADE 2

OS OBJETOS DA BIBLIOTECA DIGITAL E SUAS CARACTERÍSTICAS

2.1 OBJETIVO GERAL

Apresentar a natureza dos objetos digitais e suas características, incluindo o acesso permanente e a interoperabilidade.

2.2 OBJETIVOS ESPECÍFICOS

Esperamos que, ao final desta Unidade, você seja capaz de:

- a) identificar os elementos constitutivos da biblioteca digital;
 - b) reconhecer a natureza e as características tecnológicas dos objetos digitais;
 - c) identificar os aspectos básicos de interoperabilidade em bibliotecas digitais.
-

2.3 INTRODUÇÃO

Uma biblioteca digital possui um acervo composto por objetos digitais que têm características bem particulares, diferentes do material bibliográfico impresso. Não só as características são diferentes, mas também a diversidade desse acervo e sua origem. Por exemplo, em vez de apenas documentos textuais, a biblioteca digital pode ter um acervo de material multimídia. Esse material pode ter nascido em meio digital ou ter sido digitalizado a partir de textos ou outros tipos de materiais analógicos, como fitas cassete VHS. Seja como for, esse nosso objeto digital precisa ser preservado, tanto quanto o material impresso de uma biblioteca convencional. Para isso, precisamos entendê-lo em sua complexidade, a fim de garantir que continuará a ser usado no longo prazo.

2.4 ELEMENTOS CONSTITUTIVOS DA BIBLIOTECA DIGITAL

Os elementos que fazem parte da biblioteca digital são os recursos disponibilizados nela e o seu acervo. Além disso, percebe-se que ela é, em sua essência, um sistema com uma missão bem definida, que oferece serviços para um determinado público-alvo, ou seja, entende-se a biblioteca digital como uma iniciativa de caráter institucional. Esses aspectos lhe fornecem uma identidade que permite diferenciá-la de outras entidades semelhantes.

Os recursos que a biblioteca digital pressupõe são de duas naturezas:

- tecnológicos;
- humanos.

Utilizar recursos tecnológicos significa que é necessário ter uma infraestrutura de *software* e *hardware* para que a biblioteca digital seja implementada e mantida funcionando.

Um tipo de recurso tecnológico é o *software* (ou conjunto deles) utilizado para apoiar a realização dos serviços executados pela biblioteca digital. Além desse, ainda são necessários os *softwares* que fornecem a infraestrutura básica que se conecta com a biblioteca digital, tais como sistemas operacionais e sistemas gerenciadores de bancos de dados, entre outros.

Outro tipo de recurso tecnológico é o *hardware*, ou seja, os computadores que vão hospedar os *softwares* utilizados e o material digital, os equipamentos e materiais necessários para formar uma rede (idealmente conectada à internet, especialmente se entendemos que a biblioteca digital tem um alcance para além da instituição local). Além disso, há também os equipamentos utilizados para digitalização de



Nato digital

Diz-se do arquivo que foi criado originalmente em meio digital, em contraponto ao arquivo digital que foi criado a partir da digitalização de um material analógico (impresso ou em outro meio não digital, como uma fita VHS).



documentos, se a biblioteca digital se propuser a disponibilizar material que não é nato digital e ainda não está digitalizado. Os conceitos relativos à digitalização serão estudados na Unidade 4.

Por sua vez, os recursos humanos são responsáveis pela gestão da biblioteca digital, pela formulação de políticas, processos, aquisição de insumos, entre outras tarefas. Isso significa que é necessário pessoal capacitado para apoiar e manter o funcionamento da biblioteca digital. Considerando que ela é, em essência, um sistema de informação com as características básicas de uma biblioteca, espera-se que, entre o pessoal capacitado, haja bibliotecários. Considerando, ainda, que a biblioteca digital tem forte dependência da tecnologia, incluindo seu acervo, que é de natureza digital, espera-se também que haja pessoal da área de tecnologia da informação e comunicação (TIC). Ou seja, é preciso um grupo de pessoas, de diferentes perfis profissionais, alocados para manter a biblioteca funcionando, não só do ponto de vista de sua infraestrutura de *software* e *hardware*, como também dos diversos serviços que se constituem na finalidade da biblioteca digital, materializados no compromisso que ela assume com seu público-alvo.

A biblioteca digital, como qualquer outra, possui um acervo. Porém, no caso da digital, como mencionado anteriormente, o acervo é composto por objetos digitais, que podem ter sido criados originalmente em meio digital ou a partir de um processo de digitalização. Seja qual for sua origem, para serem facilmente recuperados e preservados, são descritos por metadados (como já explicado na Unidade 1).

Os objetos digitais podem ser simples ou complexos. Quando complexos, suas partes podem ser reutilizadas para formar outros objetos em bibliotecas digitais.

Como podemos observar, os objetos digitais são de natureza diversa, e essa natureza tem uma influência marcante nas características de uma biblioteca digital. O fato de estar em meio digital confere ao objeto algumas especificidades distintas. Elas serão abordadas a seguir, juntamente com a apresentação da natureza do objeto digital.

Objeto digital

É o mesmo que documento digital. De acordo com o CONARQ (2011), um documento digital é aquele que se encontra codificado em dígitos binários e que, para ser acessível, necessita de um sistema computacional, ou seja, *software* e *hardware*.

Objeto digital complexo

É um objeto digital composto de mais de um tipo de objeto. Todos os objetos digitais compostos possuem propriedades internas que informam como as partes devem ser montadas para formar o objeto complexo. Objetos digitais complexos podem também possuir relações externas explícitas com outros objetos digitais, tais como *hiperlinks*, que fornecem um meio de navegar de um objeto para outro (HEADSTROM; LEE, 2002).



2.4.1 Atividade

Enumere quais são os componentes básicos de uma biblioteca digital, situando o papel dos metadados.

Espera-se que você cite os componentes de *software* e *hardware*, os recursos humanos e o acervo, composto por objetos digitais. O papel dos metadados é descrever os objetos digitais para sua identificação, acesso e preservação em longo prazo.

2.5 NATUREZA E CARACTERÍSTICAS TECNOLÓGICAS DO OBJETO DIGITAL

Como sabemos, o acervo das bibliotecas digitais é constituído fundamentalmente por objetos digitais. Porém, é importante destacar que, quando se trata do mundo digital, a nomenclatura pode ser ambígua, como se observa no trecho a seguir:

Com frequência os termos arquivo de computador (*file*), documento digital, artefato digital, recurso digital e material digital são utilizados como sinônimos de Objeto Digital (YAMAOKA; GAUTHIER, 2013, p. 80).

Essa ambiguidade pode ser entendida devido às múltiplas facetas pelas quais um objeto digital pode ser enxergado, e que são importantes para compreender sua natureza.

De um lado, temos uma *entidade conceitual*, um documento que é capaz de ser interpretado e entendido por uma pessoa, e que é materializado em um suporte físico. De maneira análoga a um documento impresso, em que o suporte é o papel, no documento digital o suporte é a mídia (DVD, memória *flash*, HD etc.) e, como tal, ocupa espaço, ao contrário do que algumas pessoas pensam quando usam o termo “documento virtual”. Virtual ou não, os objetos digitais têm um *aspecto físico*, o seu suporte, que lhe confere uma materialidade.

Além disso, no meio digital os objetos estão organizados de forma *lógica*. Seu conteúdo, ou seja, sua parte conceitual precisa ser codificada e organizada.

Na codificação, o conteúdo do documento é registrado em uma sequência de números binários (ou *bytes*). De uma forma bem simplificada, podemos explicar isso pelo fato de que os computadores funcionam com circuitos eletrônicos por meio do uso de sinais, permitindo apenas representar os estados 0 (não energizado) e 1 (energizado). Por essa razão, o conteúdo do documento tem de ser codificado de acordo

Bit e byte

Cada dígito binário (0 ou 1) é denominado *bit* (do inglês, *binary digit*). Um grupo de oito *bits* é denominado de *byte*. Os computadores armazenam dados em *bytes*. Cada *byte* pode representar 256 caracteres, dependendo do formato de codificação.

Formato de codificação

É uma convenção sobre como representar caracteres (letras, sinais de pontuação, dígitos numéricos, entre outros) utilizando códigos numéricos, ou seja, quais os códigos numéricos que são associados a cada caractere.

Exemplos de formatos de codificação comuns são o ASCII (*American Standard Code for Information Interchange*) e o Unicode. De acordo com a tabela ASCII, por exemplo, a letra "A" corresponde ao número 65 em decimal; a letra "B" corresponde ao número 66 em decimal, e assim por diante. Naturalmente, ao serem representados no computador, esses números são convertidos para o código binário.

Formato de arquivo

É um modelo que determina de que maneira os *bytes* de um arquivo digital devem ser organizados de modo a armazenar seu conteúdo propriamente dito, bem como outras informações necessárias para sua correta exibição, como dados sobre formatação, tipo de fontes, metadados de autoria, tamanho etc.

Especificação de formato de arquivo

É um documento que explica de que modo o formato de arquivo deve ser estruturado. A partir dela, os desenvolvedores de *software* vão ser capazes de entender como devem criar, ler ou alterar os arquivos com o formato correspondente.

com esses sinais. Por exemplo, não podemos registrar a letra "A" como letra em si, como ocorre no documento em papel, mas, sim, como um código numérico; nesse caso, será o número "65", que, expresso em binário, seria o número "0100001". Existem diversas propostas de codificar caracteres em números, de acordo com os diversos formatos de codificação existentes.

Mas não se trata apenas de codificar os caracteres. É necessário, ainda, organizar os dados em um formato de arquivo que seja capaz de fornecer uma estrutura adequada para representar a forma de seu conteúdo, como, no caso de um texto, suas margens, negritos, sublinhados etc. Sem essa forma lógica de organização explicitada, os objetos digitais seriam apenas sequências de 0 e 1 que poderiam representar qualquer coisa. Por exemplo, uma mesma sequência de "010100011 011100011 010100000" pode significar coisas distintas: caracteres, instruções de formatação etc., dependendo do formato de arquivo. Cada formato de arquivo possui uma especificação, que pode ser de diversas naturezas.

O conteúdo e a estrutura de um objeto conceitual devem ser contidos de alguma forma no objeto lógico ou nos objetos que representam o objeto na forma digital. No entanto, o mesmo objeto conceitual pode ser representado em diferentes codificações digitais. Por exemplo, um mesmo documento, gravado e reproduzido no processador MS-Word™ e no *Adobe Portable Document Format*™ (pdf), podem manter o mesmo conteúdo, aparência e estrutura. Isto é, têm a mesma aparência visual. (YAMAOKA; GAUTHIER, 2013, p. 83).



Explicativo

As especificações de formato de arquivo podem ser abertas ou fechadas. Se abertas, significa que estão disponíveis publicamente e livres de direitos autorais (*copyright*) e, se fechadas, significa que não estão disponíveis publicamente, ou estão, mas seu uso está protegido por direitos autorais. Especificações de formato de arquivo abertas também podem ser proprietárias (quando têm um dono) ou não proprietárias (quando não o têm). As fechadas sempre têm um dono. Exemplos de formatos de arquivo fechado são: cdr, doc, xls. Já exemplos de formatos de arquivo aberto são: pdf, pdf-a, odf. O formato de arquivo pdf-a permite que "[...] a aparência visual seja preservada por um longo período de tempo, independente de ferramentas e sistemas usados para criar, salvar e reproduzir o documento" (PDF-A, 2009, p. 5).

Não confunda o fato de o formato de arquivo ser aberto com o fato de o arquivo com esse formato poder ou não ser editado. São coisas distintas. A princípio, qualquer arquivo pode ser editado, basta adquirir (ou desenvolver) o *software* adequado.

Thibodeau (2002) sumariza os diferentes aspectos que, juntos, formam a noção de um objeto digital:

Cada objeto digital é um objeto físico, um objeto lógico e um objeto conceitual, e suas propriedades em cada um desses níveis podem ser significativamente diferentes. Um objeto físico é simplesmente uma inscrição de sinais em algum meio físico. Um objeto lógico é um objeto que é reconhecido e processado pelo *software*. O objeto conceitual é o objeto que é reconhecido e entendido por uma pessoa, ou em alguns casos reconhecido e processado por um programa de computador capaz de executar transações comerciais (THIBODEAU, 2002, p. 3).

Esses diferentes aspectos são ilustrados na Figura 4, em que temos um exemplo de um documento com a foto de um cão. O documento representa um objeto conceitual que está, no caso do exemplo, no formato de arquivo jpg (objeto lógico) e armazenado em um DVD (objeto físico).

Figura 4 – Aspectos do documento digital



Fonte: adaptado de Ferreira (2006).

Rosely Rondinelli ilustra esses conceitos de forma bem compreensível:

Consideremos primeiramente o fato de que um documento digital [...] exibido numa tela de computador muda completamente a partir do momento em que o "salvamos", ou "fechamos". Isto porque, nesse momento, o documento deixa de ser compreensível aos olhos humanos para se transformar numa cadeia de *bits* (*bit strings*). Cada vez que esse documento é chamado à tela de novo, um mecanismo é acionado, no qual a cadeia de bits é processada por um *software* que a transforma no documento passível de leitura e compreensão (RONDINELLI, 2011, p. 233).

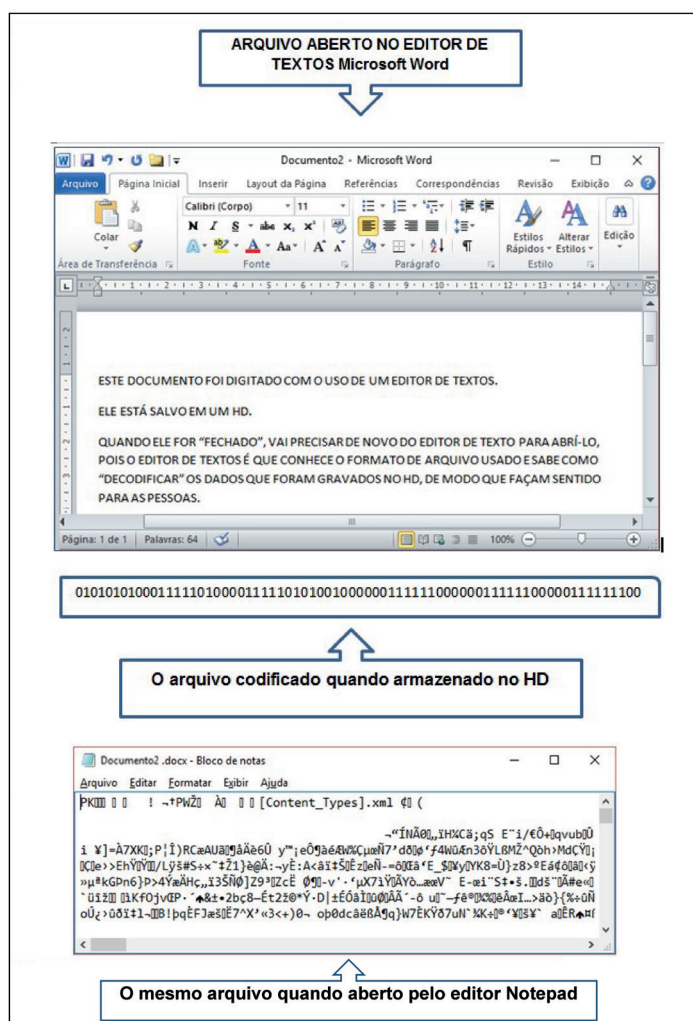
A Figura 5 ilustra essa dinâmica. Nela, podemos perceber a importância dos formatos de arquivo para a correta interpretação do conteúdo digital armazenado. Se utilizarmos um *software* que interpreta o conteúdo digital de acordo com uma especificação de formato diferente da que foi utilizada para criar o arquivo, o que será exibido em tela será diferente do que aquilo que o usuário estaria esperando. Isso se o *software* conseguir abrir o arquivo, pois ele pode reconhecer que o arquivo está em um formato incompatível com o esperado.

Dessa forma, os objetos digitais, como documentos, só podem ser lidos ou ter seu conteúdo acessado quando intermediados pela tecnologia de informação (*hardware* e *software*). E, ainda, podem apresentar uma variedade de tipos predominantes de conteúdo (texto, imagens, vídeos etc.), que vão influir em suas características específicas a serem representadas, como cores, imagens, sons, legendas etc. Esse conteúdo é codificado em *bytes*, de acordo com um formato de codificação, e eles, por sua vez, são organizados de acordo com um formato de arquivo. Além disso, como possuem uma materialidade, são registrados em algum tipo de suporte digital de armazenamento ou memória secundária, por exemplo, discos ópticos (CD, DVD), disco rígido (HD), entre outros.

Outras características dos objetos digitais os tornam peculiares:

- são editáveis (permitem alteração de modo regular e contínuo), tanto de forma bem quanto mal-intencionada;
- permitem interação com o usuário, podendo apresentar comportamentos dinâmicos associados (por exemplo, mudar sua forma de exibição ou o conteúdo, o idioma, o tamanho da fonte etc.);
- seus componentes podem estar distribuídos em diferentes locais (instituições distintas, máquinas distintas) (KALLINIKOS; AALTONEN; MARTON, 2010).

Figura 5 – O papel do *software* na decodificação do documento digital



Fonte: produção da própria autora.

Do ponto de vista da Arquivologia, é importante traçar uma distinção entre interatividade e dinamicidade. Segundo *Duranti e Thibodeau* (2006, p. 30), em um documento digital interativo, a interação do usuário pode alterar o conteúdo e a forma como a informação é mostrada. Nesse tipo de documento, as regras que determinam o conteúdo e a forma de apresentação podem ser fixas ou variáveis. Se fixas, o documento é considerado interativo, mas não dinâmico. Se variáveis, é considerado interativo e dinâmico. Um exemplo do primeiro caso é um catálogo de vendas *on-line*. Já exemplificando o segundo caso, pode-se citar uma página *web* com informações sobre o tempo. Neste último, dependendo da escolha do usuário em relação à cidade e à data, por exemplo, o conteúdo pode mudar.

Os objetos digitais apresentam muitas vantagens. Dentre elas, podemos citar:

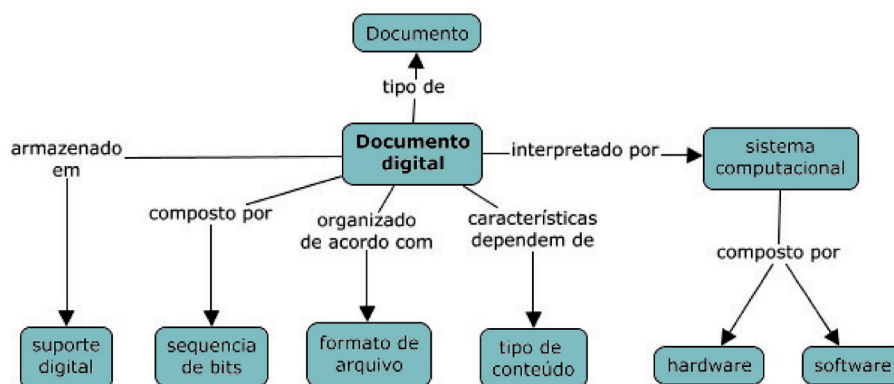
- maior rapidez em sua elaboração;
- arquivamento de forma simples e fácil recuperação de dados;
- custo de armazenamento reduzido;
- possibilidade de duplicação e transmissão imediata;
- dificuldade de fraude, mediante mecanismos que a impeçam;
- capacidade de resistência ao envelhecimento e à deterioração (GANDINI; JACOB; SALOMÃO, 2012).

Porém, eles têm também uma grande desvantagem, que é a dependência da tecnologia. A necessidade de intermediação é um dos pontos fracos do documento digital, uma vez que, para que possamos ter conhecimento de determinada informação que se encontra armazenada em forma de *bits*, precisamos de um computador e de *softwares* que sejam compatíveis.

Em resumo, observamos, então, que os objetos digitais são artefatos de natureza diversa e, além de sua materialidade, são compostos por seu conteúdo, bem como por metadados, e podem apresentar comportamento e relação com outros objetos.

A Figura 6 reúne os diferentes aspectos dos objetos digitais mencionados anteriormente.

Figura 6 – Mapa conceitual sobre documento digital



Fonte: produção da própria autora.

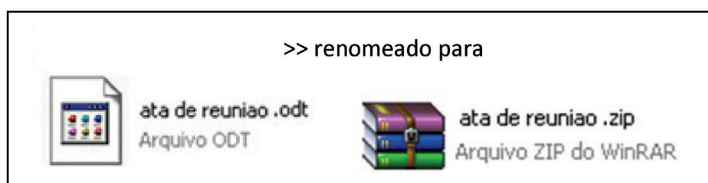


Curiosidade

O formato aberto odt é, na verdade, uma “cápsula” onde diversos dados e metadados estão armazenados de forma compactada, em formatos textuais abertos, como o xml. Faça um teste! Crie um documento no formato odt, como uma ata de reunião, e depois renomeie o arquivo gerado para a extensão .zip.

Por exemplo, se seu arquivo é denominado “ata de reunião.odt”, mude seu nome para “ata de reunião.zip”. A seguir, abra esse arquivo “ata de reunião.zip”, e veja como o arquivo “ata de reunião.odt” é, na verdade, composto por vários outros, todos encapsulados como se fossem um só.

A seguir, o conteúdo de “ata de reunião.zip” mostrado quando se abre esse arquivo:



Nome	Tamanho	Comprimido	Tipo	Modificado
..			Pasta	
Configurations2			Pasta	4/4/2012 19:35
META-INF			Pasta	4/4/2012 19:35
Pictures			Pasta	4/4/2012 19:35
Thumbnails			Pasta	4/4/2012 19:35
content.xml	47.947	4.822	XML Document	4/4/2012 19:35
layout-cache	49	46	Arquivo	4/4/2012 19:35
manifest.rdf	532	244	Arquivo rdf	4/4/2012 19:35
meta.xml	1.143	1.143	XML Document	4/4/2012 19:35
mimetype	39	39	Arquivo	4/4/2012 19:35
settings.xml	9.145	1.619	XML Document	4/4/2012 19:35
styles.xml	141.133	7.921	XML Document	4/4/2012 19:35



2.5.1 Atividade

Explique as semelhanças e diferenças básicas entre um documento digital e um documento convencional em papel, por exemplo, situando suas vantagens e desvantagens.

	Documento convencional	Documento digital
Vantagens		
Desvantagens		

Resposta comentada

O documento digital e o convencional têm, ambos, um aspecto conceitual e um suporte. No caso do documento convencional em papel, o suporte é o papel e ele não requer nenhum tipo de intervenção tecnológica para ser lido. Já no documento digital, o suporte é algum tipo de mídia ou memória secundária digital, como um DVD, CD etc. Além disso, para ser criado e lido, ele precisa da intermediação tecnológica de *software* e *hardware*. Essa intermediação implica a codificação e a decodificação do conteúdo do documento, de acordo com um formato de codificação, e, ainda, a estruturação do conteúdo codificado em um formato de arquivo. Como vantagens do documento digital, podemos citar a facilidade de criação, alteração, acesso, duplicação e armazenamento, uma vez que ele ocupa menos espaço físico. Como desvantagem, tem-se a necessidade da mediação tecnológica, que é bastante sujeita à obsolescência.

2.6 A QUESTÃO DA INTEROPERABILIDADE

Conforme afirma Sayão (2009, p. 8), a biblioteca digital é um ambiente que deve prever “[...] um complexo de serviços e de coleções de conteúdos distribuídos, gerenciados de forma autônoma, contudo interoperáveis.”

A definição de interoperabilidade fornecida é tomada no âmbito de sistemas de informação e, como tal, tem o foco nos aspectos técnicos. Entretanto, na Ciência da Informação, no contexto de bibliotecas digitais, outros aspectos são observados. Assim, considera-se a interoperabilidade como uma questão mais ampla, em que o efetivo compartilhamento de recursos armazenados em diferentes locais depende, também, de acordos entre instituições, que são de natureza política, dependendo das relações estabelecidas entre pessoas no âmbito desses locais:

O desafio da interoperabilidade é caracterizado pela multiplicidade de facetas que ela possui: interoperabilidade técnica, interoperabilidade semântica, in-

Interoperabilidade

No contexto de integração e interoperabilidade entre bibliotecas, Marcondes e Sayão (2001, p. 4) a definem como “a possibilidade de o usuário buscar recursos informacionais heterogêneos, armazenados em diferentes locais em uma rede, utilizando-se de uma interface única e sem necessidade de conhecimento sobre como os recursos estão armazenados”.

Semestre

7

teroperabilidade política e humana e muitas outras (SAYÃO, 2008, p. 27).

No que tange ao aspecto técnico, a questão está no uso de *softwares* que consigam trocar mensagens e dados sem barreiras de segurança ou outras restrições que venham a impedir essa “conversa”. Para isso, um aspecto importante é que os *softwares* utilizem protocolos adequados e compatíveis. Por vezes, tratar a questão da interoperabilidade do ponto de vista técnico não é simples, pois podem ser necessários ajustes que impliquem algum esforço de adaptação dos *softwares* e acordos para sua configuração, em relação a seus aspectos de segurança. Mas até mesmo essas questões envolvem não só a parte técnica em si, como também acordos entre as instituições envolvidas:

Assegurar a plena interoperabilidade exige frequentemente uma mudança profunda na forma pela qual uma biblioteca digital trabalha, relaciona-se com as organizações parceiras, usuários e fornecedores e, especialmente, sua atitude diante dos problemas relacionados à informação. O desafio de projetar serviços coerentes para uma diversidade de usuários a partir de componentes que são tecnicamente diferentes e gerenciados por diferentes organizações exige um sofisticado grau de cooperação (SAYÃO; MARCONDES, 2008, p. 137).

Como podemos observar, diversos são os aspectos envolvidos na interoperabilidade em bibliotecas digitais. Dentre eles, destacamos os relacionados à questão técnica: o acesso permanente aos objetos digitais e o compartilhamento desses objetos entre bibliotecas digitais. Esses tópicos serão abordados a seguir.

A interoperabilidade semântica envolve a questão dos metadados e seu significado acordado. Embora seja pertinente, ela será abordada apenas na Unidade 3, uma vez que metadados não são exclusivos para uma determinada finalidade, como promover a interoperabilidade em si: existem outros aspectos envolvidos, como a recuperação da informação e a preservação digital. A questão dos acordos organizacionais não será abordada, devido às muitas especificidades desse tópico. Detalhes sobre aspectos de acordos organizacionais podem ser obtidos em Sayão e Marcondes (2008).

2.7 A QUESTÃO DO ACESSO PERMANENTE AOS OBJETOS DIGITAIS

Como sabemos, o acervo das bibliotecas digitais é formado por objetos digitais. Esses objetos possuem, como uma de suas características, a possibilidade de apontar para outros, utilizando-se de *hiperlinks*, que

são acessíveis através da *web*. Esse tipo de objeto é um objeto digital complexo, que depende da estabilidade desses *links* para que possa ser compreendido em toda a sua completude, em especial se considerarmos que esses objetos referenciados pelos *links* podem estar distribuídos em locais diferentes do repositório onde está sediada a biblioteca digital.

Por exemplo, uma biblioteca digital no Rio de Janeiro pode ter um documento armazenado nela que faz referência a outro documento, armazenado em uma biblioteca digital no Pará. Mas o que aconteceria se esse *link*, que aponta para o documento do Pará, se perdesse, ou seja, se recebêssemos a mensagem “página não encontrada” ao clicar nele? Certamente você já deve ter passado por alguma situação semelhante. Mas por que isso ocorre? Para entender essa questão, precisamos explicar de que forma os *links* funcionam na *web*.

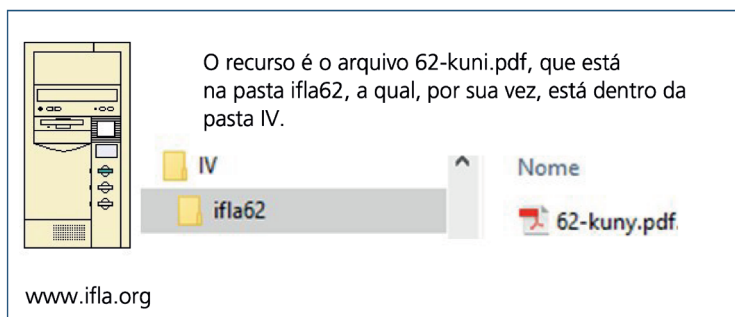
2.7.1 A nomeação das URLs

Um *link*, no contexto de um documento disponível pela *web*, é uma URL (*Uniform Resource Locator*), ou seja, um endereço lógico que serve para localizar um recurso (página, foto, vídeo etc.) fisicamente na *web*. Certamente você já digitou muitas URLs em seu navegador. Por exemplo, algo como a URL a seguir:

<<http://www.ifla.org/IV/ifla62/62-kuny.pdf>>.

Se analisarmos a estrutura de formação dessa URL, vemos que ela aponta para um recurso que é, na verdade, um arquivo pdf nomeado 62-kuny.pdf. Esse arquivo está armazenado fisicamente em uma estrutura de pastas (no caso desse exemplo, nomeada IV/ifla62) em um servidor (no caso do exemplo, associado a www.ifla.org). Por fim, existe a indicação do nome do protocolo, que, no caso, é o http. A Figura 7 ilustra esse conceito.

Figura 7 – Estrutura de uma URL



Fonte: produção da própria autora.

Essa forma de nomeação faz com que esse tipo de identificador, a URL, seja muito susceptível a mudanças em seu nome. Existem várias formas de isso acontecer, dentre elas: o pessoal de TI pode mudar o protocolo para o protocolo https (que é mais seguro) ou, então, o que é mais comum, eles podem resolver renomear ou reorganizar a estrutura de pastas (por exemplo, para IV/2016/ifla62). Ou, ainda, renomear o nome do arquivo PDF.

Basta ocorrer uma dessas situações para que o *link* mude de nome e, assim, o original deixe de funcionar. Por exemplo, se apenas o nome do arquivo for renomeado para artigo-kuny.pdf, a URL passará a ser nomeada <http://www.ifla.org/IV/ifla62/artigo-kuny.pdf>, e se alguém tentar acioná-la com o nome original, não vai encontrar o recurso, recebendo uma mensagem do tipo:

“A página requisitada “/IV/ifla62/62-kuny.pdf” não pode ser encontrada”.

Resta agora saber como podemos fazer algo a respeito, de modo a evitar que a referência para o recurso se perca, quando o nome da URL for mudado. É aí que entra o conceito de identificador persistente.

Identificador persistente

“É um nome para um recurso digital que permanece o mesmo para sempre, independente da localização do recurso” (SAYÃO, 2007, p. 68). Esse nome segue regras padronizadas, cuja lei de formação depende do tipo de solução de identificador persistente adotada.



2.7.2 Identificadores persistentes

A ideia por trás dos identificadores persistentes é simples: uma URL nomeada de forma especial é utilizada nos *links* (vamos considerar que seja uma URL equivalente), em vez de se usar a URL nomeada de acordo com a localização física do recurso (a URL verdadeira), como se faz comumente. Essa URL equivalente aponta não para o recurso em si, mas para um serviço de *software* que vai “descobrir” qual é a localização física do recurso desejado, ou seja, qual a URL verdadeira. Naturalmente, para que isso seja possível, alguém tem de informar a esse serviço de *software* qual é o nome atual da URL que aponta para a localização física atual do recurso.

Existem várias formas de implementar a resolução de nomes de identificadores persistentes para achar a URL que aponta para a localização física do recurso. Dentre elas, destacamos duas básicas: instalação, na própria instituição, de um *software* “resolvedor de *links*”, ou seja, cuja função é achar a correspondência entre o identificador persistente e a URL atual, que aponta para o recurso, ou a contratação de um serviço de *software* externo à instituição (SAYÃO, 2007). Um exemplo do primeiro caso é o PURL (*Persistent URL*) e um do segundo é o DOI (*Digital Object Identifiers*).

2.7.3 Identificadores persistentes baseados em PURL

Na solução do PURL, em vez de usarmos uma URL que aponta diretamente para a localização física do recurso, usamos uma URL persistente que aponta para um serviço de *software*, o qual, a partir do identificador persistente, consulta uma base de dados que contém a URL atualizada do recurso correspondente à URL persistente. Para isso, a URL persistente tem a seguinte lei de formação: protocolo, endereço do serviço de *software*, nome do recurso. Por exemplo, na URL persistente <http://purl.oclc.org/artigos/62-kuny.pdf>, temos que <http://purl.oclc.org> é o endereço do serviço de *software* e [artigos/62-kuny.pdf](http://purl.oclc.org/artigos/62-kuny.pdf) é a parte que identifica o nome do recurso (incluindo eventuais estruturas de pastas) (SHAFER et al., 1996). Se, por exemplo, o documento pdf (62-kuny.pdf) referenciado pela URL persistente estivesse atualmente na URL <http://meusite.com/artigos/62-kuny.pdf>, então, ao clicarmos na URL persistente, esta redirecionaria o usuário para a URL atual.

A solução do PURL é gratuita.

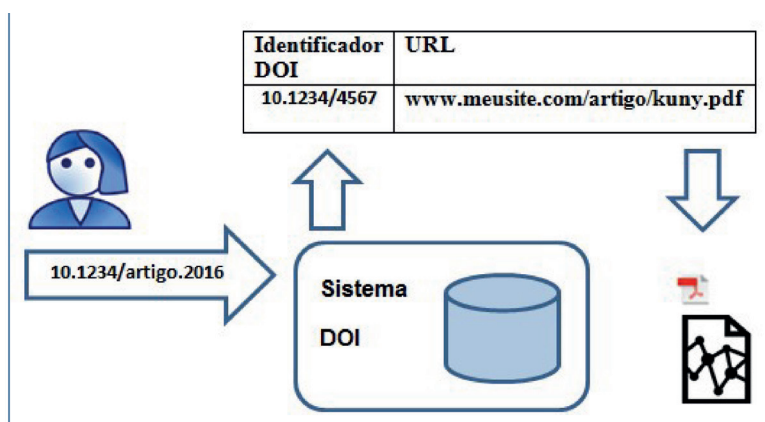
2.7.4 Identificadores persistentes baseados em DOI

A solução do DOI funciona de maneira análoga à do PURL, no sentido de haver também uma URL persistente com uma lei de formação específica e um serviço de *software* para encontrar a URL atual a partir da URL persistente. Entretanto, existem algumas diferenças básicas: a lei de formação da URL persistente é diferente, a solução é paga e não pode ser instalada em servidores locais da própria instituição, como no PURL. Além disso, a solução de nomeação do PURL para a URL persistente não provê uma infraestrutura de identificação tão robusta como a DOI (PASKIN, 2003, p. 51).

A lei de formação dos identificadores persistentes do DOI obedece a uma divisão em duas partes. A primeira consiste em um prefixo que nomeia o produtor do documento e começa com "doi:10.", seguida de um complemento para cada produtor em si. A segunda é um sufixo que identifica o documento em si. As duas partes são separadas por uma barra (/). O sufixo é gerado automaticamente pelo serviço contratado para a solução DOI, ou, então, é informado manualmente pelo usuário do serviço, respeitando-se a regra de formação (TONKIN, 2008). Um exemplo válido de identificador DOI é: doi:10.1234/artigo.2016. Nele, temos "doi:10.1234" como o prefixo que identifica o produtor e "artigo.2016" como o sufixo que identifica o documento em si. O funcionamento básico simplificado do DOI é ilustrado na Figura 8.

Na ilustração simplificada, um usuário utiliza o identificador persistente para acessar o recurso desejado (o documento pdf que está disponível em alguma URL verdadeira, que ele não sabe qual é, nem precisa saber). O sistema DOI, então, possui uma tabela, em um banco de dados, com informações de qual é a URL verdadeira atual para cada identificador persistente. Dessa forma, ao receber a solicitação do usuário para um determinado identificador persistente, basta consultar essa tabela para obter a URL atual e direcionar o pedido do usuário para essa URL. Naturalmente, para que isso funcione, alguma pessoa autorizada tem de ter acesso ao sistema DOI para atualizar o nome dessa URL sempre que ele mudar. No exemplo da figura, o usuário acessa o sistema DOI com o identificador persistente 10.1234/artigo.2016 e o sistema, então, ao consultar a tabela, direciona o pedido do usuário para a URL atual, que é <www.meusite.com/artigo/kuny.pdf>.

Figura 8 – Funcionamento simplificado do DOI

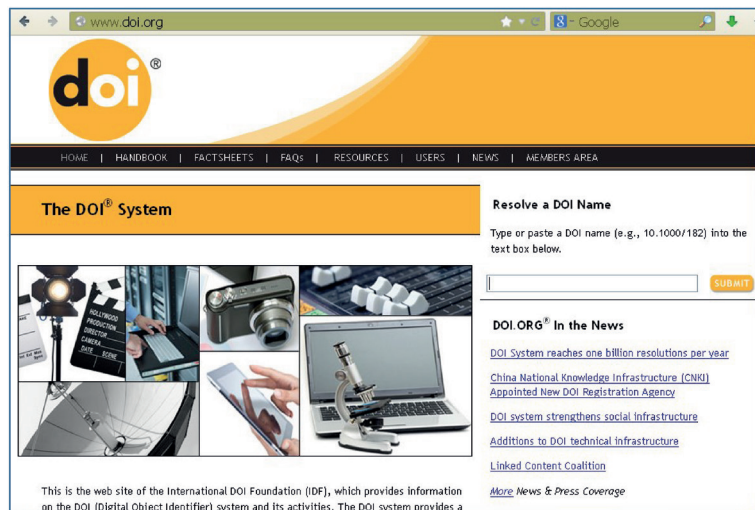


Fonte: produção da própria autora.



Multimídia

Faça uma experiência para encontrar um artigo disponível na web a partir de seu identificador persistente. Entre no site do DOI (<www.doi.org>). Você deve ver uma tela semelhante à exibida a seguir.



Depois, informe o identificador persistente, indicado na sequência, na caixa de texto à direita da tela, ao lado do botão de SUBMIT.

O identificador persistente é: 10.5433/1981-8920.2011v16n2p140.

A seguir, clique no botão de SUBMIT e veja como o DOI descobre onde o artigo está, na verdade, e encaminha o seu navegador para essa URL atual.

<<http://www.uel.br/revistas/uel/index.php/informacao/article/view/10389>>.



2.7.5 Atividade

Explique qual o objetivo dos identificadores persistentes e sua importância.

Resposta comentada

Identificadores persistentes são usados para resolver o problema dos “links quebrados”. Isso se dá pela forma como as URLs são constituídas, contendo em seu nome dados da localização física do recurso que estão sujeitos a mudar, por exemplo, o nome do recurso e as pastas onde ele se encontra. Identificadores persistentes são importantes para que um recurso esteja sempre acessível aos usuários, mesmo que mude de localização física.



2.7.6 Atividade

Assinale, na lista a seguir, os aspectos que são pertinentes para a interoperabilidade em bibliotecas digitais, além dos identificadores persistentes.

1. () uso de metadados padronizados
2. () uso de protocolos compatíveis
3. () estabelecimento de acordos institucionais
4. () adoção de medidas de segurança de rede

Resposta comentada

Itens 1 a 3.

Como podemos perceber, os identificadores persistentes desempenham um papel importante nas bibliotecas digitais, pois, se bem administrados, asseguram aos usuários o acesso permanente aos documentos digitais que estão no acervo da biblioteca, mesmo que esses objetos mudem de local dentro das máquinas que os abrigam.

Parece claro, portanto, que os *links*, entre os recursos digitais, cumprem um papel importante no estabelecimento da funcionalidade entre bibliotecas e arquivos digitais configurados de forma distribuída. Isto porque a interoperabilidade entre esses repositórios é garantida – entre outras coisas – pela confiança de que os enlaces entre os recursos por eles disponibilizados permaneçam estáveis por longo período de tempo (SAYÃO, 2007, p. 66).

Se consideramos, ainda, que as bibliotecas digitais se encontram interligadas, essa importância se acentua, cabendo às instituições parceiras garantir que esse serviço de identificadores persistentes funcione



para cada uma delas, de modo a contribuir para a relevante questão da garantia do acesso.

2.8 A QUESTÃO DO COMPARTILHAMENTO DE OBJETOS DIGITAIS

No final da década de 1990, a falta de interoperabilidade entre repositórios de artigos acadêmicos (denominados *e-prints*) deixa evidente a dificuldade nos processos de disseminação desses artigos. Dentre os problemas-chave existentes, estavam as múltiplas interfaces de busca separadas, dificultando achar os documentos desejados. Além disso, as instituições não conseguiam partilhar entre si, de forma automatizada, os metadados usados para descrever esses documentos. Nesse contexto, em julho de 1999, um grupo de pesquisadores do *Los Alamos National Laboratory*, nos Estados Unidos, convocou especialistas para uma convenção em Santa Fé, no Novo México, para tentar articular soluções para tais problemas. Essa convenção deu origem ao surgimento do **protocolo** OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*) (OAFORUM, 2016).

No contexto de uso do protocolo OAI-PMH, cada recurso possui um registro em um repositório, com metadados associados a ele. O protocolo permite a busca distribuída aberta a recursos para coleta de metadados. Sendo assim, é importante observar que o OAI-PMH é um protocolo para coleta de metadados e não para consulta de conteúdo em tempo real, como é o caso, por exemplo, do protocolo Z39-50. Seu objetivo é promover a interoperabilidade entre aplicações, facilitando o compartilhamento dos metadados. Sua última versão data de 2015.



Explicativo

O protocolo Z39-50 é utilizado para buscas distribuídas, em diversas fontes de informação. Para isso, cada fonte de informação deve ser disponibilizada por meio de um *software* que implemente esse protocolo e cada fonte a ser pesquisada deve ser configurada nas instituições que desejam fazer a busca.

Mas como funciona o protocolo OAI-PMH? Para ele funcionar, temos um conjunto de participantes, cada um com seu papel, e um conjunto de

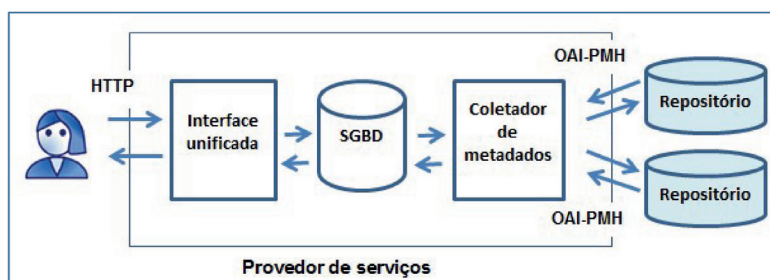
comandos fornecidos pelo protocolo, que serve para interagir com esses participantes, solicitando informações.

Os participantes da Iniciativa são divididos em provedores de dados e provedores de serviços. Ambos têm de implementar o protocolo OAI-PMH em suas soluções de *software* para repositórios. Os provedores de dados têm o papel de manter repositórios acessíveis pela internet com documentos digitais, permitindo, por meio do protocolo, o acesso aos metadados associados a esses documentos. Os provedores de serviços, por sua vez, disponibilizam serviços para busca ou coleta desses metadados dos servidores de dados. Um exemplo desse tipo de serviço seria o acesso unificado a acervos de diferentes provedores de dados, por meio de um portal *web* único (MARCONDES; SAYÃO, 2002). O SciELO (<www.scielo.org>) é um exemplo de implementação de interface com o protocolo OAI-PMH em uma base de dados já existente (MARCONDES; SAYÃO, 2003).

Outro exemplo é a Biblioteca Digital Brasileira de Teses e Dissertações do IBICT. Nessa proposta, é formada uma rede de instituições de ensino e pesquisa que atuam como provedores de dados. O IBICT opera como coletador, colhendo metadados de teses e dissertações dos provedores, fornecendo serviços de informação sobre esses metadados e expondo-os para serem coletados por outros provedores de serviços (BDTD-IBICT, 2016).

A Figura 9 ilustra esse processo, em que um usuário acessa uma interface unificada em um provedor de serviços. Essa visão unificada do usuário é possível devido ao serviço de coleta de metadados executado pelo provedor. No exemplo da Figura 9, ilustra-se uma coleta feita a partir de dois repositórios. Poderiam ser muito mais. Os dados coletados são armazenados em um banco de dados, o qual é consultado pelo sistema que provê a interface unificada.

Figura 9 – Funcionamento simplificado de um provedor de serviços usando o protocolo OAI-PMH



Fonte: produção da própria autora.

Cabe destacar que o coletador de metadados pode optar por coletar somente os metadados novos que foram acrescentados nos repositórios onde a coleta é feita, ou os que foram modificados desde a última coleta efetuada.

Os serviços oferecidos pelo protocolo OAI-PMH são viabilizados através de um conjunto de seis verbos ou comandos do protocolo.

Os verbos ou comandos são usados para fazer requisições de serviços. Essas requisições são acionadas através do mesmo protocolo usado para a *web*, o HTTP, e os comandos são os seguintes: *Identify* (recupera

informações gerais sobre um repositório, tais como seu nome e o e-mail do administrador), *ListRecord* (coleta um conjunto de registros de metadados em um repositório), *GetRecord* (dado um identificador de um registro e o prefixo do tipo de metadado de retorno, recupera os metadados individuais desse registro contido em um repositório), *ListIdentifier* (uma forma abreviada de *ListRecord*, em que são recuperados apenas os identificadores de registros, em vez dos registros em si), *ListMetadataFormats* (recupera os formatos de metadados disponíveis no repositório, como o *Dublin Core*), *ListSets* (recupera dados sobre a estrutura hierárquica de assuntos que classificam os documentos no repositório) (OAORG, 2016). As respostas a essas requisições são codificadas em XML.

Ao proporcionar uma maneira de interligar recursos dispersos em diferentes locais e soluções de *software* (desde que elas implementem o protocolo), promove-se a interoperabilidade. Uma das vantagens proporcionadas pela interoperabilidade entre bibliotecas digitais é a possibilidade de elas compartilharem um acervo ampliado, formado pela coleta de metadados das instituições parceiras interligadas.



Explicativo

Um exemplo de requisição de serviço, via HTTP, utilizando o protocolo OAI-PMH, é exibido a seguir. Nele, a requisição é usada para recuperar os metadados no padrão oai_dc, do objeto identificado por oai:dspace.c3sl.ufpr.br:1884/335, no repositório existente em <http://dspace.c3sl.ufpr.br/dspace-oai>.

A seguinte URL foi usada para fazer a requisição:

```
<http://dspace.c3sl.ufpr.br/dspace-oai/request?  
verb=GetRecord&identifier=oai:dspace.c3sl.ufpr.br:1884/335  
&metadataPrefix=oai_dc>.
```

Ao colocar tal URL no navegador, o resultado é retornado como a seguir (em uma página *web*):

```
-<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.oai.org  
<responseDate>2013-02-22T14:05:01Z</responseDate>  
<request metadataPrefix="oai_dc" verb="GetRecord" identifier="oai:dspace.c3sl.ufpr.br:1884/335">  
-<GetRecord>  
-<record>  
-<header>  
  <identifier>oai:dspace.c3sl.ufpr.br:1884/335</identifier>  
  <timestamp>2005-03-31T13:41:10Z</timestamp>  
  <setSpec>hdl_1884_286</setSpec>  
</header>  
-<metadata>  
-<oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc.xsd">  
  -<dc:title>  
    Coordenação do Chute do Futebol Sob Condições de Fadiga em Sujeitos Saudáveis</dc:title>  
  <dc:creator>Andrade, Sergio Luiz Ferreira</dc:creator>  
  <dc:description>Orientador : Andre Luiz Felix Rodacki</dc:description>  
  <dc:description>Inchi bibliografia e anexo</dc:description>  
  <dc:description>Area de concentracao: Exercicio e Esporte</dc:description>  
  -<dc:description>
```

2.9 CONCLUSÃO

Uma biblioteca digital possui elementos cuja natureza e características têm impacto direto na forma como o profissional da informação deve conduzir suas atividades no mundo de hoje. Entender esses objetos, suas características e suas origens (nato digitais ou digitalizados) é de fundamental importância, pois cada vez mais os documentos digitais ocupam espaço em bibliotecas digitais, que se constituem em um novo espaço de atuação do bibliotecário. É preciso entender o objeto digital em sua complexidade, seu contexto, saber suas especificidades e sua relação com a tecnologia, incluindo a questão da interoperabilidade.

RESUMO

A seguir, uma relação dos tópicos mais importantes que foram apresentados na Unidade 2.

Elementos constitutivos

Os elementos que fazem parte da biblioteca digital são os recursos tecnológicos e humanos que ela disponibiliza, bem como seu acervo. Este é composto por objetos digitais, que podem ser criados originalmente em meio digital (nato digitais), ou ser resultado de um processo de digitalização. O objeto digital pode ser simples ou complexo. Este último caso ocorre quando ele é composto por mais de um objeto digital simples, por exemplo, um documento digital que possui *links* para outros documentos digitais. Para ser acessado com mais facilidade e também ter seu contexto preservado, deve ser descrito por metadados.

Além disso, a biblioteca digital é um sistema com uma missão bem definida que presta serviços para um determinado público-alvo, ou seja, ela é entendida como uma iniciativa de caráter institucional.

Natureza e características tecnológicas dos objetos digitais

Um objeto digital é um documento eletrônico codificado em dígitos binários e acessado por meio de um sistema computacional. O fato de ser codificado implica a existência de um processo de representação da informação contida no documento que é materializado no computador. Para isso, o conteúdo passa pela codificação de seus caracteres (se for de natureza textual) e, ainda, pela estruturação de sua forma, o que é estabelecido em seu formato de arquivo.

Cada objeto digital possui três facetas:

- a) é um objeto físico e, como tal, possui um suporte, que é a mídia onde fica armazenado;
- b) é um objeto lógico e, como tal, é codificado e estruturado em um formato de arquivo;



c) é um objeto conceitual e, como tal, possui um conteúdo intelectual que deve ser preservado.

A questão da interoperabilidade: identificadores persistentes

As URLs são constituídas contendo, em seu nome, dados da localização física do recurso. Esses dados estão sujeitos a mudar, por exemplo, quando há alteração no nome do recurso e nas pastas onde ele se encontra. Isso faz com que, quando o recurso muda de lugar, a URL deixe de ser acionável e, portanto, os usuários deixem de ter acesso ao recurso naquele endereço. Identificadores persistentes são importantes para que um recurso esteja sempre acessível a seus usuários, mesmo que ele mude de localização física. Exemplos de implementações de identificadores persistentes são o PURL (gratuito) e o DOI (pago).

A questão da interoperabilidade: o protocolo OAI-PMH

O protocolo OAI-PMH permite a busca distribuída aberta a recursos para a coleta de metadados, fornecendo, assim, uma solução para a interoperabilidade entre bibliotecas digitais. Para ele funcionar, temos um conjunto de participantes, cada um com seu papel, e um conjunto de comandos fornecidos pelo protocolo, que serve para interagir com esses participantes, solicitando informações.

Os participantes da Iniciativa são divididos em provedores de dados e provedores de serviços. Os provedores de dados possuem o papel de manter repositórios acessíveis pela internet com documentos digitais, permitindo, por meio do protocolo, o acesso aos metadados associados a esses documentos. Os provedores de serviços, por sua vez, disponibilizam serviços para busca ou coleta desses metadados dos servidores de dados.