

A APLICAÇÃO DO DESENVOLVIMENTO E GERENCIAMENTO DE COLEÇÕES NA CONSTRUÇÃO DE REPOSITÓRIOS INSTITUCIONAIS

artigo de revisão

Rita de Cássia do Vale Caribé*

RESUMO

Consiste em revisão de literatura com o objetivo de traçar um panorama quanto à aplicação de princípios, práticas e metodologias de desenvolvimento e gerenciamento de coleções, utilizados normalmente em bibliotecas, para construção de repositórios, independentemente do seu tipo, visto que há uma carência de estudos nessa área, tanto no Brasil quanto no exterior. O estudo aborda os repositórios institucionais de maneira ampla, incluindo tanto os provedores de dados quanto os de serviços. Reflete sobre os aspectos pertinentes ao desenvolvimento de coleções, tais como a definição do escopo da coleção, a elaboração de uma política de desenvolvimento de coleções, aspectos relacionados aos direitos e propriedade intelectual, tipos de input e output dos quais deverão derivar os metadados. Apresenta parâmetros a serem considerados quanto a sustentabilidade e avaliação dos diretórios.

Palavras-chave

REPOSITÓRIO INSTITUCIONAL
DESENVOLVIMENTO DE COLEÇÃO
GESTÃO DE COLEÇÃO

* Doutoranda em Ciência da Informação pela Universidade de Brasília. Mestre em Biblioteconomia e Documentação pela Universidade de Brasília. Graduada em Biblioteconomia pela Universidade Federal de Minas Gerais.
E-mail: rita.caribe@gmail.com

I INTRODUÇÃO

De acordo com Genoni (2004) o conteúdo é um problema crucial para o sucesso dos repositórios institucionais que vêm sendo desenvolvidos por uma ou mais instituições. Assim este artigo tem como objetivo traçar um panorama quanto à aplicação de princípios, práticas e metodologias de desenvolvimento e gerenciamento de coleções, utilizados normalmente em bibliotecas, para construção de repositórios, independentemente do seu tipo. Para isso foi elaborada uma revisão de literatura, sendo consultadas bases de dados especializadas tais como: *Library and Information Science Abstracts (LISA)*, *Library Literature and Information Science Full Text* e *Library and Information Science and Technology Abstracts*.

As mudanças decorrentes do desenvolvimento tecnológico têm provocado

alterações no campo social e econômico, como também nos valores culturais e sociais, que envolvem a tecnologia, gerando como consequência diversas inovações. Sandler (2006) aponta um termo criado por sociólogos - "*tiping point*" - para caracterizar quando movimentos quantitativos ou mudanças incrementais ocorrem, ultrapassando as suas fronteiras e produzindo como consequência mudanças qualitativas.

A criação do computador pessoal, da *Web*, a emergência do comércio eletrônico que possibilita um amplo acesso a bens e serviços, a hegemonia da cultura digital como característica definidora de uma geração globalizada, são tipos de mudanças que necessariamente afetam as unidades de informação.

Bodi e Maier-O'Shea (2005) apresentam três características da pós-modernidade, que interferem diretamente nas unidades de informação, a saber:

a) Resistência a hierarquias e controle

Os bibliotecários definem regras e políticas que são utilizadas para dar suporte à decisão quanto ao descarte de um objeto físico que não consideram mais necessário. Têm por tradição coletar, identificar e organizar fisicamente recursos informacionais de acordo com esquemas de classificação, o que resulta em pouca ou nenhuma flexibilidade para tratar recursos interdisciplinares. Por outro lado, o meio eletrônico foi idealmente adaptado para prover múltiplos pontos de acesso para materiais interdisciplinares, porém o controle torna-se dificultado. No meio eletrônico os materiais são muitas vezes efêmeros, deixam de estar acessíveis sem nenhum aviso prévio; não há controle sobre quais os periódicos que continuam nas bases eletrônicas de texto completo e quais os que apareceram subitamente. Uma instituição que efetua uma assinatura de uma base de dados não tem controle sobre todos os títulos incluídos nessa base, não mais possui a coleção física sobre a qual tinha controle, fazia empréstimos ou preservava.

b) Pensamento não linear

O mundo ocidental tem tradicionalmente o pensamento linear, enfatiza o raciocínio lógico e a importância da narrativa. Por outro lado, o ambiente *online* não é linear, é amorfo, com *links* que conduzem a inesperadas direções, a idéias não esperadas, a descoberta de padrões inimagináveis e em alguns casos, a não encontrar nada. O hipertexto não está subordinado ao pensamento linear, nem a categorias, nem a sistemas de classificação, que por sua vez não são adequados para pesquisa, raciocínio e compreensão em um modo interdisciplinar. Os textos não são mais fixos, são fluidos, encontrando padrões de conexão. Nesse novo contexto é preciso repensar o desenvolvimento de coleções e do acesso bibliográfico.

c) Fronteiras indefinidas

Há uma indefinição de fronteiras que é expressa por meio de textos interdisciplinares, pelas mudanças de formas e formatos no ambiente eletrônico, tornando-se difícil distinguir entre periódicos eletrônicos, bases de dados *full-text*, índices, catálogos, *Web sites* e multimídias.

As mudanças sejam elas lentas ou abruptas são processos naturais, assim as bibliotecas de hoje, em decorrência, principalmente das

mudanças no campo da informação, deverão ser fundamentalmente diferentes daquelas de 20 anos atrás. Alguns pesquisadores vêem isso como uma revolução, outros como um processo natural e constante. As bibliotecas vêm acompanhando continuamente e se adaptando ao longo dos séculos (SANDLER, 2006).

Os profissionais de informação deram, durante muito tempo, demasiada importância ao suporte, à informação registrada. Atualmente há uma diversidade de mídias que promovem a aprendizagem e a transferência da informação tanto quanto o livro, e as bibliotecas devem e estão incorporando essas outras mídias. Sandler (2006), considera perda de tempo continuar a discutir a importância do livro como mídia, mesmo porque o livro faz parte da história do mundo, se constituiu ao longo do tempo na raiz do trabalho do profissional bibliotecário e porque nos próximos anos, e ainda por um período razoável de tempo, os livros ainda serão publicados, consumidos e continuarão a interferir na vida dos indivíduos.

Galbreath (1997 *apud* BODI; MAIER-O'SHEA, 2005, p. 143) afirma que uma coleção deve ser vista como "[...] *a fundamentally intellectual construct rather than exclusively as an assembly of physical objects [...]*". Neste mesmo sentido Sandler (2006, p.240) afirma que "[...] *libraries are not about books; they were, are, and will be about facilitating communication across time and space. Books have been a way to do that historically, but today there are other, often better, ways do accomplish this [...]*".

Complementando essa linha de raciocínio Bodi e Maier-O'Shea (2005) afirmam que uma coleção não está mais limitada a uma coleção física em um único local, mas a uma mistura de local e remoto, de papel e eletrônico. Nesse sentido questionam quais as melhores formas de prover acesso: papel ou eletrônico, local ou remoto, proprietário, empréstimo, licença ou compartilhamento por meio de consórcios. A tendência atual é de coleções híbridas.

2 DESENVOLVIMENTO DE COLEÇÕES

Bibliotecas existem há aproximadamente 5.000 anos, desde então sempre houve um processo, formal ou informal, composto por critérios, estratégias e metas, para coletar objetos de informação que seriam incorporados

às coleções dessas bibliotecas. Inicialmente denominou-se esse processo de seleção, num segundo momento de desenvolvimento de coleções (*collection development*) e mais recentemente de gerenciamento de coleções (*collection management*). De maneira geral, se constitui em um processo universal, a ser desenvolvido por qualquer tipo de unidade de informação -- biblioteca, centro de informação etc. -- ou instituição que esteja interessada em fornecer algum tipo de serviço ou produto de informação. Deverá necessariamente adotar um conjunto de critérios, estratégias e metas para coleta, seleção, descarte etc. dos objetos de informação que pretende incorporar à sua coleção, de forma dinâmica e constante (EVANS; SAPONARO, 2005).

Desenvolvimento de coleções ou gerenciamento de coleção, termo mais adotado a partir da década de 80, consiste na criação e aplicação de políticas, na tomada de decisões relacionadas direta ou indiretamente com a coleção, a avaliação do uso, a gestão dos recursos compartilhados, a cooperação, a avaliação do programa e a gestão fiscal. Integra o ciclo da informação, é influenciado pelos demais integrantes deste ciclo como também os influencia, de forma que qualquer alteração ou inclusão de novas metodologias e tecnologias no ciclo da informação deverá impactar o gerenciamento de coleções. (JOHNSON, 2005)

De maneira geral, de acordo com Evans e Saponaro (2005) e Johnson (2004), no processo de desenvolvimento e gerenciamento da coleção estão incluídas atividades como: estudo e identificação da comunidade a ser atendida e especificação do público alvo; estudo do acervo existente identificando seus pontos fortes e fracos; definição das metas de desenvolvimento do acervo; estudo e planejamento quanto à conservação e preservação do acervo; definição dos tipos de materiais que serão incluídos no acervo; definição dos tipos e critérios para parcerias, consórcios e compartilhamento de recursos uma vez que não poderá adquirir tudo o que for necessário; atenção para os problemas legais, de direito autoral, licenças de uso etc.; preocupação com aspectos relacionados à censura, à liberdade de expressão e intelectual e com minorias sociais; avaliação do acervo; definição de critérios de descarte e manutenção de duplicatas.

Os critérios gerais adotados para todo e qualquer tipo de material em uma unidade de informação são semelhantes, incluindo os recursos eletrônicos, que Johnson (2004) denomina como conceito guarda-chuva para todos os recursos digitais, os quais possuem muitos gêneros, formatos e mídias tanto de armazenamento quanto de entrega.

3 REPOSITÓRIOS INSTITUCIONAIS

Na década de 1990 a comunidade científica iniciou um movimento para o desenvolvimento e implantação de arquivos abertos, o que vem revolucionando a comunicação científica. Sob a liderança do físico Paul Ginsparg, do Laboratório de Los Alamos, dos Estados Unidos, foi desenvolvido o primeiro arquivo eletrônico de *preprints* ou *eprints*, o ArXiv. Os pesquisadores criaram um repositório de acesso livre para as áreas de física, matemática e ciência da computação, que tinha como objetivo oferecer uma alternativa para a comunicação científica na qual o próprio pesquisador pudesse publicar sua produção e que o seu trabalho ficasse armazenado em uma Biblioteca Digital, de forma que os leitores pudessem conhecer as pesquisas e submeter comentários àquele trabalho publicado. A partir de então foram criados vários outros repositórios especializados em uma área ou subárea temática como física, matemática, ciências da computação, biologia quantitativa dentre outros. Esses repositórios temáticos foram criados como uma tentativa de utilizar o poder da Internet para prover uma forma alternativa e mais barata para o acesso à literatura.

Lynch (2003) afirma que no final de 2002 algo extraordinário aconteceu em continuidade à revolução da informação, o surgimento dos repositórios institucionais como uma nova estratégia para permitir mudanças no processo da comunicação científica. Diversas tendências, esforços e desenvolvimentos tecnológicos permitiram esse avanço.

Existem alguns conceitos para repositório institucional:

- Crow (2002) apresenta que, em um sentido mais geral, um repositório institucional é um repositório digital, que pode ser de qualquer tipo de coleção de material digital, de propriedade, hospedado, controlado ou disseminado por uma faculdade ou

uma universidade, independentemente de propósito ou origem. Em um sentido mais restrito apresenta que os repositórios institucionais são arquivos digitais da produção intelectual criados por uma faculdade, grupo de pesquisa e estudantes da instituição e acessíveis a todos os usuários, internos e externos à instituição, com poucas ou nenhum tipo de barreira.

- Lynch (2003) os define como um conjunto de serviços que uma universidade oferece aos membros de sua comunidade para o gerenciamento e disseminação de materiais digitais criados pela instituição e por seus membros. É essencialmente um compromisso da instituição de zelar por esse material digital, incluindo a preservação de longo prazo quando apropriado, bem como a organização, acesso ou distribuição. Afirma que o desenvolvimento dos repositórios institucionais emergiu como uma nova estratégia para possibilitar a comunicação na comunidade científica.
- Barton e Waters (2005) apresentam o repositório institucional como uma base de dados que possui um conjunto de serviços para captura, armazenamento, indexação, preservação, redistribuição da pesquisa à comunidade científica em formato digital. De acordo com *The Scholarly Publishing & Academic Resources Coalition* (SPARC) um repositório institucional deve ter as seguintes características: ser definido por uma instituição de cunho acadêmico e científico, que tenha caráter acumulativo e perpétuo, e que utilize os conceitos de arquivo aberto e interoperabilidade, especificamente o *Open Archives Initiative* (OAI).

Embora os autores acima se refiram aos repositórios institucionais como somente de conteúdos relativos às instituições de ensino superior e pesquisa, o próprio Lynch (2003) e também Johnson (2005), fazendo um comentário ao SPARC, que admite somente o cunho científico e acadêmico, contrapõem essas afirmações conjecturando que futuramente os repositórios poderão ser utilizados por qualquer instituição que desenvolva pesquisa ou outro tipo de propriedade intelectual, que poderão também

estabelecer seus repositórios. Essas instituições podem ser departamentos de governo ou agências, organizações não governamentais ou intra-governamentais, museus, organizações independentes de pesquisa, federações, sociedades, e pelo menos teoricamente as entidades comerciais. Enfim, qualquer instituição que tenha interesse em capturar e disseminar ampla e abertamente seu produto intelectual, contribuindo desta forma com o discurso científico e acadêmico e beneficiando-se ao aumentar a visibilidade de sua organização.

De acordo com Lynch (2003) um repositório institucional não é simplesmente um conjunto de softwares e hardwares, mas também um conjunto de compromissos, decisões e atividades para administrar os materiais digitais, incluindo a sua preservação de longo prazo, a organização e o acesso ou distribuição, bem como o gerenciamento de mudanças tecnológicas e a migração de conteúdos digitais de um conjunto de tecnologias para outra, dentre outros.

Nas bases pesquisadas citadas anteriormente foram identificados apenas três artigos que tratam do desenvolvimento de coleções aplicado a repositórios institucionais: o artigo de Lynch (2003), o de Hunter e Day (2005) e o de Genoni (2004).

Hunter e Day (2005) abordam o desenvolvimento e gerenciamento de coleções aplicado aos repositórios institucionais, na perspectiva tanto dos repositórios provedores de dados e quanto dos repositórios provedores de serviços. Apresentam conceitos e práticas relativos à atividade, enfatizam a importância de uma política, bem como a necessidade de desenvolver e gerenciar as coleções de maneira cooperativa. Argumentam que, de acordo com o incremento da disponibilidade de informação digital o trabalho do bibliotecário, em alguns anos, será similar ao dos *gatekeepers* e provedores de informação digital, o qual poderá ser gerenciado de qualquer lugar.

Genoni (2004) também aborda o gerenciamento da coleção para repositórios institucionais focando o papel das bibliotecas nesse novo contexto e incita-as a tomar seu lugar na prestação de serviços digitais e a participar do processo de comunicação científica. Discute os problemas relativos aos conteúdos dos repositórios institucionais, para os quais sugere a aplicação de técnicas e metodologias já conhecidas

pelos bibliotecários para o gerenciamento da coleção.

Os repositórios institucionais são coleções que possuem seu próprio conjunto de critérios e estratégias para o desenvolvimento de coleções. Lynch (2003) argumenta que diante da nova realidade dos repositórios institucionais as bibliotecas devem estabelecer novas estratégias de desenvolvimento de coleções para o mundo digital, assumindo inclusive a responsabilidade pela administração dos conteúdos que terão importância futura. Enfatiza que os repositórios são locais onde podem ser colocados muito mais materiais do que as bibliotecas de instituições de pesquisa são capazes de identificar como coleção de valor.

As instituições responsáveis pelos repositórios, de acordo com Lynch (2003), devem ter um compromisso institucional de serem guardiões dos objetos digitais, incluindo a preservação de longo prazo quando apropriado, bem como a sua organização, acesso e distribuição. O desenvolvimento de coleções neste contexto deve significar cooperação com outras bibliotecas de pesquisa para o licenciamento de conteúdos digitais, para prover acesso ao material antigo mesmo que a assinatura tenha sido cancelada.

Hunter e Day (2005) afirmam que a arquitetura do OAI-PMH¹ (*Open Archives Initiative Protocol for Metadata Harvesting*) significa em alguns casos que o desenvolvimento e gerenciamento de coleções deve estar direcionado para dois níveis distintos em termos do que o protocolo define como provedor de dados ou provedor de serviços. Por exemplo, o provedor de dados deverá considerar os aspectos do desenvolvimento e gerenciamento de coleções relacionados com o próprio serviço, isto é, com a definição do escopo do serviço, com os níveis de qualidade. O provedor de serviços, por sua vez, deverá considerar outros tipos de problema relacionados com a informação disponibilizada por meio dos provedores de dados, tais como a comparação entre áreas temáticas cobertas, o tipo de recurso incluído e o nível de qualidade dos conteúdos selecionados.

¹ Protocolo modelo de interoperabilidade baseado no processo de coleta automática de metadados (*metadata harvesting*). Este protocolo opera sobre o protocolo http. Os provedores de serviço (PS) enviam solicitações de metadados aos provedores de dados (PD). Estes respondem com metadados estruturados em registros XML, obedecendo a um padrão de metadados. (IBICT. <http://bdtd.ibict.br/bdtd/glossario/glossario.jsp?op=1>, acesso em 05 de junho de 2007)

4 PROVEDORES DE DADOS

No contexto do Protocolo OAI-PMH o provedor de dados é entendido como a instituição que administra o sistema que mantém o repositório de dados e suporta o protocolo como forma de expor os metadados, ou melhor, mantém um ou mais repositórios (*web servers*) que suportam o OAI-PMH como forma de expor os metadados. (<http://www.openarchives.org/documents/FAQ.html>) A construção de um repositório institucional não consiste apenas da escolha de software, hardware e na solicitação para que os indivíduos da organização participem. É necessário um entendimento comum entre os integrantes da comunidade quanto aos propósitos do repositório, bem como de um conjunto de políticas que definam o escopo, juntamente com os problemas como depósito, acesso e sustentabilidade. Esses aspectos podem ser vistos como aspectos do gerenciamento ou desenvolvimento da coleção.

De acordo com Lynch (2003) e Hunter e Day (2005) existem diversos softwares, dentre eles destacam:

- o DSpace² que consiste em um software que possui diversas alternativas e possibilidades de utilização para diferentes tipos de materiais. O DSpace foi desenvolvido sob a liderança do Massachusetts Institute of Technology (MIT), que em parceria com a Hewlett Packard Corporation criou um modelo direcionando o caminho a ser seguido por outras universidades. Em 2003, com o financiamento do Andrew W. Mellon Foundation e outras fontes o DSpace foi estruturado para ser replicado em um número maior de instituições em todo o mundo; o software também foi liberado publicamente de acordo com o protocolo de arquivos abertos, a um custo mais baixo e menos barreiras para o desenvolvimento de repositórios em todas as instituições. (LYNCH, 2003)
- o software da *University of Southampton* no U.K. (<http://www.eprints.org/>), embora possua um propósito bem geral, foi projetado mais especificamente para repositórios institucionais ou repositórios

² <http://www.dspace.org/>

de documentos de disciplinas, ao contrário de materiais digitais arbitrários.

4.1 Definição do escopo da coleção

De acordo com as informações apresentadas anteriormente, o repositório institucional é um arquivo digital da produção intelectual criado por uma faculdade, grupo de pesquisa e estudantes de uma instituição (CROW, 2002). Diante deste conceito várias escolhas precisam ser adotadas com relação ao conteúdo do repositório. Genoni (2004) afirma que o conteúdo será crucial para o sucesso dos repositórios, estando ele em uma universidade ou em qualquer instituição. O problema se refere especificamente ao tipo de material que será incluído, quem será responsável pela seleção desses materiais e como essas responsabilidades se articulam com as responsabilidades já previstas para a instituição.

O primeiro ponto a ser discutido se refere à definição dos grupos de usuários para os quais o repositório se destina, quais são suas necessidades e prioridades (GENONI, 2004). De acordo com Hunter e Day (2005) essas definições deverão influenciar os tipos de *outputs* que o repositório deverá suportar, que por sua vez irão requerer diferentes tipos de metadados ou acessos.

Com relação ao escopo dos repositórios Hunter e Day (2005) enfatizam que inicialmente foram desenvolvidos esforços no sentido de utilizá-los como dados de pesquisa, de recursos de aprendizagem e de registros administrativos da instituição, porém sugerem que também podem ser incluídos documentos relativos às atividades da organização, aos registros de eventos e da vida intelectual da instituição.

De acordo com Johnson (2005) e Crow (2002) um repositório institucional pode conter qualquer produto resultante dos trabalhos realizados pelos estudantes, pesquisadores que fazem parte ou não da instituição, bem como dos demais indivíduos que trabalham na mesma, tudo dependerá das metas definidas. Podem incluir pastas eletrônicas dos estudantes, materiais didáticos utilizados em sala de aula, relatórios anuais da instituição, gravações de vídeos, programas de computador, conjuntos de dados, fotografias, e trabalhos de arte, enfim qualquer material digital que a instituição tenha o desejo de conservar. Crow (2002) acrescenta que podem conter também *pre-prints*, trabalhos

em andamento, artigos revisados pelos pares, monografias, materiais de ensino, conjuntos de dados e outros materiais auxiliares de pesquisa, artigos de conferências, teses e dissertações eletrônicas e documentos não convencionais (literatura cinza - *gray literature*).

Genoni (2004) apresenta uma lista de materiais que podem ser incluídos como *pre-prints*: documentos de trabalhos, teses e dissertações, relatórios técnicos e de pesquisas, boletins e *newsletters* de centros de pesquisa e outros documentos, documentos relacionados a bolsas de pesquisa, relatórios de pesquisas em andamento para agências de financiamento, relatórios, atas e memorandos de comitês, relatórios de pesquisa, documentação técnica e levantamentos, bem como as diversas formas de material não convencional (*gray literature*).

Barton e Waters (2005) apresentam um manual para elaboração de repositórios onde enfatizam a importância da definição dos serviços a serem oferecidos. Considera necessário responder: Qual a missão do serviço? Que tipo de conteúdo será aceito? Quem serão os usuários chave? Quem serão os *stakeholders*? Que tipos de serviços serão oferecidos, caso conte com recursos limitados? O que pode oferecer? Os serviços serão pagos? Quais serão os serviços prioritários? Quais serão as prioridades de curto e longo prazo? Quem irá prover os metadados? Haverá revisão pelos pares?

Lynch (2003) acredita que um repositório institucional maduro e plenamente compreendido poderá conter os trabalhos intelectuais da faculdade e dos alunos -- tanto pesquisa como materiais de ensino -- e também documentos das atividades da instituição, na forma de registros de eventos e apresentações, e das atividades intelectuais do cotidiano da instituição. Deverão também abrigar dados de experiências e observações coletados por membros da instituição que dão suporte as suas atividades acadêmicas. Poderá também incluir versões digitais de artigos de periódicos e de monografias acadêmicas no meio digital. Em pesquisa realizada em 2005, em 13 países, Westrienen e Lynch (2005) apresentam que do universo pesquisado, nos repositórios institucionais acadêmicos havia inserido: artigos, livros, teses, dados primários, vídeos, música e material similar, material didático e outros tipos de materiais.

Outra decisão importante se refere à revisão pelos pares, alguns repositórios aceitam somente

artigos revisados, outros apenas informam que artigos foram revisados, e disponibilizam os demais acrescentando que os mesmos não foram submetidos à revisão. Decisões semelhantes se referem a disponibilização da versão rascunho (*draft*) ou somente a versão final produzida pela entidade publicadora (HUNTER; DAY, 2005).

Lynch (2003) considera que a importância dos repositórios institucionais está subestimada quando os caracterizam como instrumentos para reestruturar a economia da publicação acadêmica, em vez de um veículo para avançar, apoiar e legitimar um espectro mais amplo para novas comunicações acadêmicas. Considera que políticas complexas, embaraçosas para admitir materiais nos repositórios institucionais, principalmente aquelas que emulam das práticas tradicionais de publicação acadêmica, tais como o uso de revisão pelos pares, são altamente contra-produtivas. Ressalta que os repositórios institucionais não são um periódico, nem uma coleção de periódicos, e não devem ser administrados como um desses, pois a comunidade universitária deve ter credencial suficiente para inserir seus materiais no repositório institucional. Argumenta, ainda, que as instituições devem estar seguras de que existem recursos práticos e restrições que cada instituição terá que elaborar. Por outro lado, ressalta que há instituições que possuem conjuntos de dados verdadeiramente grandes ou coleções multimídias que podem ser difíceis de inserir nos repositórios.

Uma decisão prática, porém importante, se refere ao formato que podem ser submetidos ao repositório, ou seja, poderá aceitar formatos do tipo HTML, PDF, PostScript, TeX ou TIFF. Cabe ressaltar, no entanto, que de acordo com os diferentes tipos de formatos serão necessários coletar metadados para definí-los (HUNTER; DAY, 2005).

Ainda com relação ao escopo da coleção, Genoni (2004) apresenta algumas perguntas que necessariamente deverão ser respondidas, tais como: o repositório deverá incluir apenas material de pesquisa ou também material de ensino? qual o número de versões (*drafts*) que poderão ser armazenadas? o repositório terá algum tipo de relacionamento com outros serviços de arquivos disponíveis na instituição? o repositório acolherá duplicatas, como e em que situação? que categorias de usuários serão atendidas, apenas o corpo acadêmico ou incluirá também pessoal administrativo?

4.2 Processo de submissão de objetos digitais

Para controlar e administrar o acesso a esses conteúdos é necessária a definição de políticas e mecanismos incluindo o gerenciamento do conteúdo e um sistema de controle das versões dos documentos. A estrutura da política do repositório e a infra-estrutura técnica devem fornecer aos gerentes institucionais flexibilidade para controlar quem pode contribuir, aprovar, acessar e atualizar conteúdos digitais oriundos de várias comunidades institucionais e grupos de interesse, incluindo departamentos acadêmicos, bibliotecas, centros e laboratórios de pesquisa e autores individuais. Vários sistemas de repositórios existentes e em desenvolvimento possuem infra-estrutura e capacidade técnica para embargar ou suspender o acesso a submissões até que o seu conteúdo tenha sido aprovado por um revisor designado. A natureza e extensão dessa revisão irão refletir as políticas e as necessidades de cada instituição individualmente, muitas vezes essa revisão simplesmente validará a autorização institucional e/ou afiliação do autor disponibilizar um material no repositório; em outros casos, a revisão será mais qualitativa e extensa, servindo como uma certificação primária (CROW, 2002).

Crow (2002) apresenta que a submissão pode ser feita pelo próprio indivíduo interessado por meio do auto-arquivamento (*author self-archiving*), que é um termo amplo aplicado para o arquivamento eletrônico, sem a mediação do editor. Na prática, esse auto-arquivamento inclui tanto a disponibilização de *pre-print* (nas diversas instâncias) e a publicação de textos por pesquisadores individuais, em *web sites* pessoais e a inclusão de sua pesquisa em servidores de *e-print* específicos por disciplinas. As políticas de editores tradicionais variam quanto ao auto-arquivamento, destacando casos em que há uma grande severidade, no que resulta em indiferença por parte dos pesquisadores acadêmicos.

A prática do processo de submissão dependerá diretamente do software adotado, embora a instituição possa desenvolver o seu próprio sistema de submissão. Geralmente os softwares possuem interface de submissão baseado na *Web*, com a possibilidade de fazer o *upload* do documento e acrescentar os metadados descritivos.

Uma vez ingressados os dados, esses são transferidos para uma área de *buffer* onde serão submetidos a um controle de qualidade, a uma análise antes de serem disponibilizados no repositório. Esse controle de qualidade poderá variar de instituição para instituição, porém alguns tipos de controles básicos são: se o indivíduo que submeteu o artigo tem autoridade para isso; se o artigo está de acordo com o escopo da instituição e se a qualidade dos metadados associados está adequada.

Hunter e Day (2005) ressaltam que a criação manual de metadados descritivos pode consumir tempo, e que essa perda de tempo pode não ser a prioridade dos pesquisadores, assim cada instituição deverá definir a necessidade de utilizar bibliotecários para auxiliar nesse trabalho, revisando e complementando os metadados. Sugerem a utilização de padrões como o que foi disponibilizado pela OCLC - entrada de autoria, classificação - e os cabeçalhos de assunto da *Library of Congress*.

Quanto à possibilidade de descarte de documentos da coleção, ou melhor, à necessidade de remover algum documento do repositório, Hunter e Day (2005) afirmam que alguns repositórios, como o ArXiv, autorizam o indivíduo que submeteu um documento a retirá-lo, mediante uma justificativa ou esclarecimentos, porém, uma versão prévia do mesmo continuará disponível (<http://arxiv.org/help/withdraw>). Remover um artigo do repositório é o mesmo que retirar um artigo de um periódico. A natureza do arquivamento dos repositórios institucionais significa que as políticas de remoção são mais plausíveis de serem mais restritivas do que repositórios por área temática, embora a remoção do acesso pode ser um problema diferente.

Crow (2002) argumenta que uma vez submetido o material ele não pode ser mais removido, exceto nos casos que envolvam ilegalidades - difamação, plágio, infringência da lei do direito autoral ou ciência de baixa qualidade (*bad science*). Será necessário definir políticas e fluxos para dar suporte à remoção ou ao recolhimento de documentos, como por exemplo, no caso de conduta imprópria de cientistas. Por outro lado a questão da responsabilidade legal nos casos de difamação e outras formas ilegais que não são aprovadas pelos repositórios.

4.3 Direito e propriedade intelectual

Um problema legal para o qual deve-se estar atento é o direito à propriedade intelectual. Esse tema foi extensivamente pesquisado, de acordo com Hunter e Day (2005), pelo projeto denominado RoMEO³ (*Rights Metadata for Open Archiving*), desenvolvido pelo FAIR Programme⁴.

Um outro problema se refere à autorização por parte da editora do periódico, caso o autor já tenha lhe transferido o direito de *copyright*. Em muitos casos as editoras autorizam a publicação do artigo em PDF no *website* do próprio autor desde que não tenha fins lucrativos, a utilizar cópia como material didático em cursos e a reutilizar figuras e tabelas. Excluem, porém totalmente a possibilidade de inclusão em *websites* de arquivos abertos, proíbe a inclusão no *site* da instituição à qual o pesquisador pertence, bem como não autorizam disponibilizar o artigo em repositórios institucionais. Outras, porém, autorizam o autor a disponibilizar uma cópia no repositório institucional seis meses após a sua publicação, com a condição de que seja feito um link para o *site* do periódico.

Um problema crucial para os repositórios é ter a garantia de que não estão intencionalmente infringindo o *copyright* ou outro direito de propriedade intelectual. Como garantia, os repositórios solicitam que o indivíduo que está submetendo um documento confirme que possui os direitos para inseri-lo; outros solicitam, como uma parte do processo de submissão do artigo, a assinatura de um tipo de acordo com os autores.

Em relação ao problema do *copyright* as instituições ou fundos que financiam as pesquisas estão tentando reafirmar os seus direitos. Algumas universidades começam a inserir o direito da instituição incorporando, em estatutos universitários e contratos de emprego, os direitos intelectuais de propriedade especialmente

3 RoMEO: projeto desenvolvido na Universidade de Loughborough (Inglaterra) e concluído em agosto de 2002. Esse projeto pesquisou os requisitos legais das diferentes partes envolvidas no depósito e disponibilização de e-prints utilizando o OAI. Gerou soluções de direitos de metadados que está sendo desenvolvido numa solução genérica de direitos OAI nos Estados Unidos. Uma lista de políticas de auto-arquivamento foi elaborada, fornecendo informação valiosa nas quais os editores autorizam o depósito de ativos, normalmente artigos de periódicos, onde podem ter tipo os direitos autorais a eles designados. Essa lista foi assimilada e disponibilizada pelo projeto SHERPA. (http://eprints.rclis.org/archive/00001596/02/FAIR_Assignation_article.pdf)

4 FAIR (Focus on Access to Institutional Resources) Programme, lançado em agosto de 2002 e com prazo de conclusão até outubro de 2005

com referência a patentes ou recursos de aprendizagem. Bide (2002 apud HUNTER; DAY 2005) afirma que a questão da posse dos direitos de propriedade intelectual se constitui em um dos problemas mais controvertidos do ensino superior, pois nos termos da Lei sobre Direitos Autorais, Projetos e Patentes do Reino Unido, de 1988, o resultado do trabalho desenvolvido na instituição pertence à instituição. No entanto, na prática, as instituições acadêmicas não exercitam este direito com relação aos direitos autorais em artigos de periódicos ou livros. Por outro lado, Lynch (2003) adverte contra universidades que poderão fazer uso dos repositórios institucionais como um meio de controlar ou assumir o direito de propriedade do trabalho intelectual que tem sido tradicionalmente controlado pelos acadêmicos.

Com relação às agências de financiamento de pesquisas, essas não podem reivindicar os direitos autorais sobre as publicações resultantes dos projetos desenvolvidos com os recursos fornecidos. Entretanto, algumas agências expressaram o desejo de exigir garantias de que os documentos gerados serão disponibilizados por meio do acesso aberto.

4.4 Acesso

Crow (2002) apresenta que os repositórios institucionais devem estar acessíveis para os usuários finais, tanto internos quanto externos à instituição, com poucas ou nenhuma barreira para o seu acesso. Enquanto isto está plenamente em conformidade com os princípios do acesso aberto e é desejável para a maioria de relatórios de pesquisa, há um número de razões pelas quais os repositórios institucionais não podem tornar todos os conteúdos disponíveis publicamente.

Conteúdos potencialmente restritos podem estar incluídos nos relatórios de pesquisa que podem ser retirados, nos relatórios ou teses que contêm informações comercialmente sensíveis, nos conjuntos de dados que estão em processo de análise, nos recursos de aprendizagem ou nos registros administrativos. Crow (2002) argumenta que alguns desses materiais podem ser distribuídos na íntegra ou em parte, podem ser distribuídos dentro da instituição, porém o conteúdo e seus metadados descritivos não devem estar normalmente disponíveis para usuários externos à organização.

Genoni (2004) defende que devem permanecer restritos documentos que por alguma razão causem algum tipo de ofensa a grupos sociais ou culturais, por outro lado, versões preliminares de relatórios e documentos de pesquisa podem permanecer restritas até que o trabalho esteja totalmente concluído. Neste caso os repositórios podem necessitar da definição de níveis de acesso diferenciados para diferentes tipos de conteúdos, ou seja, deve inserir mecanismos de controle de acesso para determinados tipos de conteúdos que são restritos, enquanto deve assegurar que os metadados descritivos estejam disponíveis para aqueles conteúdos que podem ser compartilhados de forma mais ampla.

4.5 Sustentabilidade

Um problema do desenvolvimento e gerenciamento da coleção se refere à sustentabilidade de longo prazo. Quanto a isso há dois aspectos principais a serem considerados: a necessidade de apoio financeiro, estratégico e continuado da instituição que mantém o repositório; e a necessidade de assegurar acesso, de longo prazo e de forma continuada, ao seu conteúdo.

Lynch (2003) afirma que é extremamente importante que as instituições reconheçam que o desenvolvimento de um repositório institucional é um compromisso sério, de longa duração, em relação à comunidade universitária, ao mundo acadêmico e ao público em geral, por esse motivo a decisão de desenvolver um repositório não deve ser tomada sem analisar todas as variáveis com consistência. Adverte que os repositórios podem fracassar por diversas razões, como, por exemplo, pela falta de apoio estratégico ou financeiro das instituições, ou por falha no gerenciamento ou ainda por problemas técnicos.

Por outro lado, Hunter e Day (2005) discutem que as instituições não são tão estáveis quanto aparentam ser. Na realidade diversas instituições, departamentos ou centros de pesquisa são criados, extintos, incorporados a outros ou transferidos para uma outra instituição. Embora as instituições de ensino aparentem ter uma certa estabilidade, pois raramente são extintas ou fechadas completamente, há casos em que são incorporadas ou fundidas a outros departamentos ou faculdades. Devido a essas

possibilidades os repositórios institucionais devem procurar obter apoio político do mais alto nível dentro da instituição, devem procurar desenvolver planos de contingência que podem ser implementados em caso de mudanças não esperadas. O que pode significar acordos com outras instituições ou outros repositórios como bibliotecas nacionais, como exemplos citam os casos da Elsevier Science, da Kluwer Academic, da Blackwell, da BioMed Central, da Oxford University Press e da Taylor & Francis que assinaram acordos de depósito com a Biblioteca Nacional dos Países Baixos.

Um repositório institucional tem um duplo papel perante a universidade e a comunidade acadêmica, de ser cumulativo e garantir perpetuidade, o que de acordo com Crow (2002) apresenta duas implicações. Primeiro, os critérios de submissão de conteúdo, onde um item submetido ao repositório não pode ser mais retirado, exceto nos casos já citados anteriormente. Presume-se que todos os conteúdos estarão perpetua e universalmente acessíveis, mas isso não é a realidade, pois nem tudo estará necessariamente acessível para a perpetuidade. Genoni (2004) argumenta que devido ao caráter de perpetuidade de um repositório dificilmente ocorrerá descarte de elementos que tenham sido inseridos. Crow (2002) contesta afirmando que as instituições deverão estabelecer critérios e políticas, como também desenvolver e implementar sistemas gerenciais para autorização de acesso aos conteúdos dos repositórios, tanto para atendimento interno quanto externo. Quanto à natureza cumulativa dos repositórios deve-se considerar que será necessário contar com uma infra-estrutura escalonada, pois no processamento inicial os requisitos de armazenamento e processamento são mais modestos, porém devem-se prever sistemas que possam acomodar milhares de submissões por ano, e eventualmente poder conservar milhões de objetos digitais e muitos *terabytes* de dados.

Os repositórios institucionais têm como objetivo preservar e tornar acessível os conteúdos digitais numa base de longo prazo, assim, a preservação digital e o acesso de longo prazo estão inexoravelmente ligados, um fica sem sentido sem o outro. Assegurar o acesso de longo prazo ao conteúdo dos repositórios é um outro desafio que necessita ser encarado pelos repositórios institucionais. Lynch (2005)

apresenta que as instituições devem fazer escolhas baseadas num equilíbrio de exigência da comunidade universitária e avaliações locais sobre a factibilidade técnica, as quais resultarão em listas de formatos de arquivo que eles se comprometerão em conservar em formatos acessíveis (presumivelmente por meio da migração de formatos); em outros casos, eles podem conservar os bits que compõem um arquivo, mas não oferecerá nenhuma garantia de que estes bits poderão ser interpretados no futuro sem o desenvolvimento de programas especializados para lê-los. Estas escolhas então podem ser consolidadas buscando o consenso mais amplo entre comunidade do ensino superior e a comunidade de pesquisa como uma forma de desenvolvimento de padrões que beneficiarão o trabalho ativo dos administradores e o envolvimento continuado da universidade.

Crow (2002) afirma que até a data em que publicou o trabalho não havia padrões universalmente aceitos para assegurar a longevidade e a preservação de formatos digitais, apesar de haver um conjunto de boas práticas, fluidas e em evolução, que estejam guiando o planejamento da preservação digital. A conservação de longo prazo de objetos digitais exige gerência pró-ativa e recursos consideráveis. Muitos bibliotecários permanecem céticos quanto à idoneidade dos editores para esta tarefa, devido principalmente à sua perspectiva de curto prazo. Os repositórios institucionais, no contexto do modelo de desagregação da publicação acadêmica devem manter a responsabilidade pela preservação de materiais de pesquisa nas mãos de bibliotecários, pois são profissionalmente preparados e comprometidos com isso.

O *Joint Information Systems Committee* (JISC) realizou um estudo de viabilidade de preservação de *e-print*, onde destacou a importância de formatos de arquivo, metadados e estratégias organizacionais (JAMES et al., 2003; PINFIELD; JAMES, 2003 apud HUNTER; DAY, 2005). Repositórios existentes de *e-print* frequentemente só aceitam um número limitado de formatos de arquivo, normalmente com base nas preferências de *download* dos usuários. Incluem uma mistura de softwares proprietário e abertos como, por exemplo, HTML, PDF, Post-script, TeX, MS Word, MS PowerPoint e TIFF.

James et al. (2003 apud HUNTER; DAY, 2005) recomendam que os repositórios devem

avaliar os riscos de preservação de formatos de arquivo em suas coleções e considerar um formato de conversão, isto é, baseado em padrões abertos ou XML. De acordo com Day (2004 apud HUNTER; DAY, 2005) há também a necessidade de manter a informação sobre os formatos dos arquivos armazenados, bem como considera a necessidade de um metadado apropriado para dar suporte ao processo de preservação digital.

James et al. (2003 apud HUNTER; DAY, 2005) recomendam que os repositórios devem colaborar na produção de um conjunto comum de metadados de preservação. Progressos em ambos problemas serão de importância, mas a preservação de longo prazo do conteúdo dos repositórios dependerá do desenvolvimento de estratégias organizacionais apropriadas. Sugerem que os repositórios institucionais devem eventualmente vir a se tornar repositórios digitais de segurança. (*trusted*)

Um grupo de trabalho patrocinado pelo *Research Library Group* (RLG) e pelo *Online Computer Library Center* (OCLC) definiu alguns atributos principais de repositórios digitais de segurança. De forma bem sucinta, isso inclui a necessidade de aceitar a responsabilidade pela preservação do conteúdo, de obter o controle sobre conteúdo de forma a ser capaz de preservá-lo, de demonstrar possibilidade de sustentabilidade financeira e viabilidade organizacional, de assegurar a existência de políticas documentadas e procedimentos que possam ser monitorados e avaliados, e para aderir aos padrões e às melhores práticas (RESEARCH LIBRARY GROUP, 2002 apud HUNTER; DAY, 2005).

Muitos dos repositórios existentes teriam dificuldade em cumprir todos estes critérios. Em vez disso, James et al. (2003 apud HUNTER; DAY, 2005) comentam que muitos repositórios de *e-print* existentes estão com o foco voltado para o acesso mais do que para a preservação e que o financiamento do projeto base pode não ser o melhor meio de construção da sustentabilidade de longo prazo. A melhor forma para construir os repositórios é a cooperação com especialistas em preservação ou contratação de serviços de preservação.

4.6 Avaliação

Repositórios institucionais como provedores de dados também necessitarão

fornecer amplo espectro de serviços para suas comunidades de usuários. Estes podem incluir, por exemplo, armazenamento de dados, criação e melhoramento de metadados, pesquisa e recuperação, exportação de metadados sobre resultados da pesquisa para identificar procedimentos de avaliação. Como tal, os repositórios necessitarão de avaliação periódica para assegurar que estão cumprindo seus requisitos institucionais básicos e atendendo às necessidades dos usuários (HUNTER; DAY, 2005).

Genoni (2004) discute que é necessário definir qual o padrão de qualidade que se deseja alcançar. Caso o repositório decida incluir materiais provenientes de diferentes níveis de autoridade, deverá escolher metadados que tornem essa decisão transparente para os usuários. Como por exemplo, se considerar a revisão pelos pares como padrão de qualidade, os usuários poderão escolher se desejam ou não materiais revisados pelos pares.

Com relação à avaliação Genoni (2004) afirma que há diversos tipos: Avaliação com o objetivo verificar se o que foi realizado, se os serviços que estão sendo prestados estão de acordo com as metas definidas. Avaliação quanto à quantidade do material que está inserido no repositório, se corresponde ao total produzido pela instituição, nível de cobertura em relação à totalidade que deveria abranger, avaliações similares àquelas desenvolvidas em bibliotecas para verificar a completeza da coleção e o nível de profundidade do acervo.

Quanto aos usuários faz-se necessário saber quais os grupos que estão utilizando o repositório, qual o nível de satisfação dos usuários, tanto no que se refere ao conteúdo disponibilizado quanto ao *software* e metadados utilizados.

5 PROVEDOR DE SERVIÇOS

Os provedores de serviço são aquelas entidades que utilizam os metadados coletados automaticamente dos provedores de dados, via protocolo OAI-PMH, como base para oferecer produtos e serviços de valor agregado, ou melhor, o provedor publica solicitações de dados por meio do OAI-PMH aos provedores de dados e utiliza os metadados como base para construir serviços de valor agregado (<http://www.openarchives.org/documents/FAQ.html>).

Alguns provedores de serviço concentram-se em agregar conteúdos de metadados de múltiplos provedores de dados. Por exemplo, o serviço experimental ARC da *Old Dominion University* coleta (*harvesting*) metadados de mais de 80 fornecedores de dados e os armazena em uma base de dados relacional (<http://arc.cs.odu.edu/>). O projeto OAIster da Universidade Michigan coleta (*harvesting*) metadados descrevendo um vasto leque de conteúdo digital de diversas instituições diferentes (<http://www.oaister.org/>). Outros provedores de serviço têm como foco o fornecimento de serviços de valor agregado de acordo com o tipo de funcionalidade, como exemplo apresenta a experiência do serviço experimental da Citebase desenvolvido pela *University of Southampton* que extrai dados referenciais do texto completo e os combina com metadados coletados para construir uma base de dados de citação (<http://citebase.eprints.org/>). O projeto de *e-Prints* do Reino Unido desenvolveu um provedor de serviço experimental com o objetivo de agregar e promover o acesso aos repositórios localizados no Reino Unido e testou alguns serviços de valor agregado adicionando metadados e citações (<http://eprints-uk.rdn.ac.uk/search/>).

As políticas de desenvolvimento e gerenciamento de coleções para os provedores de serviço terão algumas similaridades com as políticas dos provedores de dados, contudo, sua implementação será em grande parte definida pela quantidade e qualidade dos metadados a serem disponibilizados pelos repositórios. Entretanto, cabe ressaltar que os repositórios institucionais podem também ser provedores de serviço, ou seja, agregar metadados sobre o conteúdo de um número de provedores de dados sob seu controle. Por outro lado, provedores de serviço podem também ser provedores de dados, na medida em que disponibilizam os dados coletados (*harvest*) por intermédio do protocolo OAI.

5.1 Definição do escopo da coleção

Os serviços que agregam conteúdo a partir de múltiplos provedores de dados têm que resolver alguns dos mesmos problemas de desenvolvimento e gerenciamento de coleções que têm os repositórios. Além de selecionar de quais provedores os metadados serão coletados, os provedores de serviços necessitam tomar

algumas decisões tais como quais os tipos, nível de qualidade, área temática, origem geográfica dos conteúdos que serão disponibilizados por meio do serviço. No entanto, é necessário que haja metadados disponíveis suficientes para apoiar estas escolhas. Por exemplo, um provedor de serviços que deseja coletar principalmente metadados de resultados de relatórios de pesquisa que foram revisados pelos pares dependerá de que essa revisão tenha ocorrido no repositório e que os relatórios estejam separados dos demais que não foram revisados, e que essas informações possam ser claramente identificadas nos metadados. Princípios semelhantes podem ser aplicados caso o provedor de serviços só esteja interessado em coletar metadados sobre pesquisas, teses ou recursos de aprendizagem (HUNTER; DAY, 2005).

A definição da cobertura temática é potencialmente mais problemática. Pode ser possível selecionar conteúdos com base em sua origem, por departamentos acadêmicos ou por meio de uma análise dos termos fornecidos em campos de assunto. Uma opção alternativa seria coletar o texto completo, onde estiver disponível, e submeter a algum tipo de processo de classificação automática (HUNTER; DAY, 2005).

Outro problema que pode ocorrer está relacionado à possibilidade de duplicação de conteúdo, pois muitos relatórios de pesquisa são submetidos a múltiplos repositórios - onde originam e em mais de uma instituição. Será tarefa do provedor de serviços determinar quanto desse tipo de problema será transferido para os usuários dos serviços, e quando necessário deverá adotar procedimentos para retirar ou evitar duplicações. É possível que ocorra problemas semelhantes com versões diferentes do mesmo conteúdo.

Finalmente, o provedor de serviço não tem nenhum controle direto sobre conteúdo, assim as estratégias de desenvolvimento e gerenciamento de coleções dos provedores de serviço devem levar em consideração que alguns repositórios podem se reestruturar ou fechar, como também podem permitir que indivíduos que submeteram documentos possam retirá-los.

5.2 Metadata harvesting

A quantidade e qualidade dos metadados coletados é crucial para os provedores de

serviços. Neste ponto específico, os provedores de serviços são quase totalmente dependentes dos repositórios provedores de dados de onde os metadados são extraídos. Infelizmente, estudos de uso de metadados em contextos OAI sugerem que a qualidade varia muito (BARTON; CURRIER; HEY, 2003; HALBERT; KACZMAREK; HAGEDORN, 2003; WARD, 2003 apud HUNTER; DAY, 2005), leva ao que Halbert (2003 apud HUNTER; DAY, 2005) denomina de colisões entre formatos de metadados e problemas com controle de autoridade e duplicação.

Garantir a qualidade dos metadados é o papel fundamental dos repositórios provedores de dados. Isso irá depender da previsão dos fluxos para submissão de artigos, interfaces e controle de qualidade. Torna-se necessário contar com o apoio de profissionais de informação e da biblioteca, embora isso encareça o desenvolvimento do repositório. No entanto, num sistema de repositório genuinamente distribuído, os metadados descritivos de cada documento só serão criados uma única vez, o que tem várias vantagens sobre os custos em relação aos trabalhos práticos de catalogação de uma biblioteca.

Guy, Powell e Day (2004 apud HUNTER; DAY 2005) apresentam algumas sugestões práticas que podem ajudar a melhorar a criação de metadados de boa qualidade. Sugerem o estabelecimento de diretrizes para os conteúdos dos metadados (POWELL; DAY; CLIFF, 2003 apud HUNTER; DAY, 2005), o aperfeiçoamento dos instrumentos de criação de metadados e a implementação de processos apropriados de controle de qualidade.

Como repositórios em si, os provedores de serviço também podem ter um interesse na geração automática, na captura ou no aprimoramento dos metadados. Estes interessados na preservação digital já devem ter experiências com ferramentas que podem capturar metadados técnicos sobre formatos, mas a captura de metadados descritivos diretamente do texto completo, principalmente no que se refere a conteúdo é algo que ainda precisa ser mais estudado no contexto do OAI. O uso de formatos de textos estruturados como os baseados em XML pode ajudar, bem como os softwares baseados no *Document Object Model* (DOM) desenvolvido pelo *World Wide Web Consortium* (W3C) (<http://www.w3.org/DOM/>). (HUNTER; DAY, 2005)

O projeto de *ePrints* do Reino Unido fez uso do serviço de terceiros (utilizando a tecnologia de serviços da *Web*) que poderia ser utilizado para melhorar a coleta dos metadados pelos provedores de serviços. O primeiro destes foi o serviço *name authority* desenvolvido pelo *OCLC Research* que valida as formas de nomes de pessoas e de instituições utilizados, melhorando a consistência do uso dos nomes. O segundo, também baseado na tecnologia desenvolvida pelo *OCLC Research*, foi para um serviço que designa automaticamente os termos de classificação de assunto, ajudando assim a assegurar a consistência dos metadados de assunto através do provedor de serviços e possibilitando que a coleta de conteúdo possa ser mais acuradamente distribuída para serviços mais focados no assunto. Um terceiro serviço, fornecido pela Universidade de Southampton, analisa as referências bibliográficas nos textos de conteúdo completo, facilitando assim os *links* de citação e as análises infométricas. Todos estes serviços têm o potencial de adicionar valor às funções que podem ser oferecidas tanto pelos repositórios quanto pelos provedores de serviços. Há, no entanto, necessidade de testar essas ferramentas provenientes de serviços de terceiros e testar efetivamente seu uso em coleta realísticas (HUNTER; DAY, 2005).

5.3 Direito e propriedade intelectual

Enquanto os repositórios institucionais, como provedores de dados, estão preocupados com os direitos de propriedade intelectual fixados em ambos repositórios de conteúdo e de metadados, os provedores de serviço estão principalmente preocupados com os metadados. Isto inclui ambos os metadados que foram coletados dos provedores de dados e as complementações que foram feitas durante o processo de fornecimento do serviço (MCCLELLAND et al., 2002 apud HUNTER; DAY, 2005).

Gadd, Oppenheim e Probets (2004a apud HUNTER; DAY, 2005) observaram que a Lei do Reino Unido sobre os direitos autorais não é clara quanto aos metadados, embora registros de proteções individuais possam ser obtidos como para uma compilação e coleções de metadados, como uma base de dados. Eles concluem que, para evitar infringir o direito da base de dados, os provedores de serviço que desejam utilizar um metadado de um provedor de dados deve

procurar solicitar permissão para proceder à extração dos dados.

O levantamento do Projeto RoMEO mostrou que a maioria dos provedores de dados concordaram que coleções de metadados gozavam dos mesmos direitos de uma base de dados e sentiram que isto deveria estar implícito nos contextos do OAI. Os provedores de serviço estão divididos entre aqueles que verificam o estado dos direitos dos metadados antes de os coletar, aqueles que consideram que sob a aplicação do protocolo OAI estes são implicitamente gratuitos, e aqueles que nunca pensaram no assunto. A maioria dos repositórios estava satisfeita por ver outros coletarem seus metadados, embora algumas condições de uso sejam especificadas.

Hunter e Day (2005) consideram que talvez haja uma necessidade de que os provedores de serviços definam suas próprias hipóteses de trabalho sobre direitos de propriedade intelectual com referência à coleta de metadados (ou texto completo) a partir de repositórios.

5.4 Sustentabilidade

Os provedores de serviços não têm com o que se preocupar em relação à gerência de longo prazo e administração por não terem diretamente os conteúdos, por outro lado têm uma necessidade de estabilidade organizacional e financeira, conforme afirmam Hunter e Day (2005). Os provedores de serviço fazem a intermediação entre provedores de conteúdo e seus usuários e, como tal, sua sustentabilidade dependerá de como fornecem o serviço que está sendo exigido e com que qualidade o executam, tendo em vista que estão num mercado potencialmente competitivo. Como serviços, no entanto, eles podem desenvolver-se ao longo do tempo, ambos em relação ao conteúdo que coletam e os serviços de valor agregado que fornecem.

Muitas iniciativas de provedores de serviço são financiadas numa base de projeto, serviços de demonstração de valor agregado que podem ser fornecidos por metadados coletados utilizando o protocolo OAI. No futuro, será possível contar com uma economia mista de provedores de serviços, com algum suporte financeiro do setor público enquanto outros serão fornecidos numa base comercial ou semicomercial (HUNTER; DAY, 2005).

5.5 Avaliação

Como serviços focados nos usuários os provedores de serviço necessitarão de avaliação permanente para assegurar que continuam a satisfazer as necessidades dos usuários e a ajudar na identificação de novas solicitações ou demandas. Os resultados de tal avaliação podem levar a mudanças nas políticas de desenvolvimento e gerenciamento de coleção (HUNTER; DAY, 2005).

6 CONCLUSÃO

Conforme podemos constatar, alguns problemas pertinentes ao desenvolvimento e gerenciamento de coleções foram identificados, os quais deverão ser considerados por ocasião da construção de repositórios institucionais. Apesar do desenvolvimento e gerenciamento de coleção, em grande parte, estar preocupado com conteúdo, como em qualquer outra biblioteca digital, decisões necessitam ser tomadas quanto aos tipos de objetos a serem incluídos ou excluídos dos repositórios. Enquanto o foco primário de um repositório é definido pelo seu alcance institucional, decisões adicionais necessitam ser tomadas sobre o tipo de conteúdo, se conterà artigos revisados pelos pares, resultados de pesquisas, teses e dissertações, recursos de aprendizagem etc.

É recomendado por Hunter e Day (2005) que os repositórios institucionais (como fornecedores de dados) elaborem formalmente suas políticas de desenvolvimento e gerenciamento de coleções, que devem conter o alcance pretendido pela coleção com referência à abrangência temática, ao tipo de objeto e à qualidade. A política também pode incluir informações sobre os fluxos para submissão de trabalhos, incluindo informações sobre os direitos de propriedade intelectual, acesso e sustentabilidade do repositório. Como exemplo apresenta o guia para indivíduos que estejam interessados em depositar objetos digitais no repositório produzido pelo *University College London*, que inclui informação quanto à elegibilidade de materiais a serem incluídos, autorizações relativas aos direitos autorais, formatos autorizados, criação de metadados e procedimentos para *upload*. (<http://eprints.ucl.ac.uk/DepositGuide.htm>).

APPLICATION OF THE COLLECTION DEVELOPMENT AND MANAGEMENT IN THE CONSTRUCTION OF INSTITUTIONAL REPOSITORIES

ABSTRACT

This work gives a broad view of the collection development and management as to the application of principles, practices, and methodologies frequently used by libraries to the construction of repositories independently of their kind once there seems to exist a lack of studies in such area, both in Brazil and abroad. This study approaches all institutional repositories in a wide range, including data providers as well as service providers. It also reflects on the pertinent aspects to the collection development, such as the definition of the collection scope, the creation of a collection development policy, aspects related to rights and intellectual property and the types of input and output from where the metadata will be derived. It presents parameters to be considered regarding sustainability and directories evaluation.

Keywords

INSTITUTIONAL REPOSITORY
COLLECTION DEVELOPMENT
COLLECTION MANAGEMENT

Artigo recebido em 13.11.2007 e aceito para publicação em 16.01.2008

REFERÊNCIAS

ATKINSON, Ross. Six key challenges for the future of collection development: introduction for the Janus Breakout Sessions. *Library Resources & Technical Services*, v. 50, n. 4, p. 244-251, Oct. 2006.

BARTON, Mary R.; WATERS, Margaret M. *Creating an institutional repository*: LEADIRS Workbook. 2005. 134p. Disponível em: <<http://www.dspace.org/implement/leadirs.pdf>>. Acesso em: 20 abr. 2007.

BODI, Sonia; MAIER-O'SHEA, Katie. The Library of Babel: making sense of collection management in a postmodern world. *Journal of Academic Librarianship*, v.31, n.2, p. 143-150, Mar., 2005.

CROW, Raym. *The case for institutional repositories*: a SPARC position paper. 2002. 37p. Disponível em: <www.arl.org/sparc/bm~doc/ir_final_release_102.pdf>. Acesso em: 20 abr. 2007

EVANS, G. Edward; SAPONARO, Margaret Zarnosky. *Developing library and information center collections*. 5.ed. Westport: Libraries Unlimited, c2005. 446p.

GENONI, Paul. Content in institucional repositories: a collection management issue. *Library Management*, v. 25, n. 6-7, p.300-306, 2004.

HUNTER, Philip; DAY, Michael. *Institutional repositories, aggregator services and collection development*. 2005. Disponível em: <<http://eprints-uk.rdn.ac.uk/project/docs/studies/coll-development/coll-development.pdf>>. Acesso em: 2 abr. 2007.

JOHNSON, Peggy. *Fundamentals of collection development and management*. Chicago: ALA, 2004. 342p.

JOHNSON, Richard K. Institutional repositories: partnering with faculty to enhance scholarly communication. *D-Lib Magazine*, v.8, n.11, Nov. 2002. Disponível em: <<http://www.dlib.org/dlib/november02/johnson/11johnson.html>>. Acesso em: 29 abr. 2007.

LYNCH, Clifford A. Institutional repositories: essential infrastructure for scholarship in the digital age. *ARL*, n. 226, p. 1-7, Feb., 2003. Disponível em: <<http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>>. Acesso em: 05 jun. 2007.

OPEN Archives Initiative. Disponível em: <<http://www.openarchives.org/documents/FAQ.html>>. Acesso em: 07 jun. 2007.

PHILLIPS, Linda L.; WILLIAMS, Sara R. Collection development embraces the digital age: a review of the literature 1997-2003. *Library*

Resources & Technical Services, v. 48, n. 4, p. 273-299, Oct. 2004.

SANDLER, Mark. Collection development in the age day of Google. *Library Resources & Technical Services*, v. 50, n. 4, p. 239-243, Oct. 2006.

WESTRIENEN, Gerar van.; LYNCH, Clifford A. Academic institucional repositories: deployment status in 13 Nations as of Mid 2005. *D-Lib Magazine*, v. 11, n.9, Sep. 2005. Disponível em: <<http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>>. Acesso em: 15 maio 2007.