COOPERAÇÃO INTERINSTITUCIONAL NUM REPOSITÓRIO DE BITS COMPARTILHADO[1]

Eld Zierau
The Royal Library of Denmark, PO Box 2149, DK – 1016 Copenhagen K E-mail: elzi@kb.dk

Ulla Bøgvad Kejser
The Royal Library of Denmark, PO Box 2149, DK – 1016 Copenhagen K E-mail: ubk@kb.dk

Resumo

Este artigo explora o modo pelo qual instituições independentes, como arquivos e bibliotecas, podem cooperar para gerenciar um repositório de bits, com preservação de bits, com o objetivo de usar os seus recursos para preservação de forma mais eficiente. O Modelo de Referência OAIS é usado para fornecer uma estrutura para uma análise sistemática dos requisitos técnicos e organizacionais das instituições, para um repositório de bits remoto. Em vez de visualizar um repositório de bits simplesmente como um Arquivamento para os repositórios das instituições, nós defendemos vê-lo como um subconjunto de funções de todas as entidades definidas pelo Modelo de Referência OAIS. O trabalho é motivado por, e foi usado em um estudo de viabilidade dinamarquês para a criação de um repositório nacional de bits. O estudo revelou que, dependendo de seus objetivos e das coleções que guardam, as instituições têm requisitos variáveis, como para segurança dos bits, acessibilidade e confidencialidade. Este estudo revelou ainda que requisitos para o nível de segurança dos bits devem ser complementados por análise de risco, a qual deve envolver elementos da arquitetura; por exemplo, o numero de cópias e como é garantida a independência entre elas. O artigo descreve a arquitetura do repositório de bits e as vantagens em ser flexível, a fim de oferecer serviços diferenciados com respeito a, entre outras coisas, segurança dos bits e custos. Além disso, são mostrados os desafios na formulação de vários aspectos, tais como requisitos de risco.

[1] NOTA DOS TRADUTORES (NT): Traduzido para o português por Miguel Rio Branco Nabuco de Gouvêa e revisado por Rubens Ribeiro Gonçalves da Silva, com a autorização das Autoras. Cf. o original em ZIERAU, Eld; KEJSER, Ulla B. *Cross-Institutional Cooperation on a Shared Bit-Repository*. **World Digital Libraries**: An International Journal, v.6, n.1, p. 25-36, June, 2013. ISSN (impresso): 0974-567X; ISSN (online): 0975-7597. Original disponível em: https://content.iospress.com/articles/world-digital-libraries-an-international-journal/wdl120098. DOI: 10.3233/WDL-120098. Acesso em: 05 set. 2018.

1 INTRODUÇÃO[2]

Existe uma conscientização crescente de que a operação e gerenciamento de sistemas de preservação digital requerem recursos contínuos e conhecimento altamente especializado. Como resultado, as instituições buscam meios de terceirizar ou compartilhar a responsabilidade sobre essas atividades, de forma a utilizar seus recursos para preservação digital mais eficientemente, e potencialmente se beneficiar de uma economia de escala. Uma área óbvia a ser explorada neste contexto é a preservação de bits, porque os requisitos para assegurar que os bits permaneçam intactos e acessíveis são relativamente bem entendidos, comparados à preservação funcional (lógica), isto é, assegurando que os bits permaneçam compreensíveis e utilizáveis. Na prática, já existem organizações oferecendo serviços de preservação de bits, como, por exemplo, a OCLC Digital Archive[3] e a Iron Mountain.[4] Outros exemplos de organizações e soluções para preservação de bits, mas também alguma preservação funcional, são a Kopal[5] e o LOCKSS.[6]

Dentro dessa linha, o Ministério da Cultura Dinamarquês financiou um projeto para investigar a viabilidade de um repositório de bits compartilhado, para os arquivos nacionais, bibliotecas e museus na Dinamarca. O projeto é conduzido pelo The Danish National Archives, The Royal Library e The State and University Library, que estão previstas para serem as partes interessadas e envolvidas (*stakeholders*) no repositório de bits, junto com outras instituições com necessidades de preservação de longo prazo. A meta é projetar um sistema comum que possibilite meios seguros e em grande escala para admitir/inserir (*ingesting*), armazenar,

---

[2] NT: Cf. o vídeo que apresenta este texto como uma de suas fontes, apresentado no Japão em setembro de 2017, cujo original está disponível em: https://vimeo.com/233024801. Acesso em: 03 set. 2018. Há versão desse vídeo traduzida e dublada em português disponível em: https://vimeo.com/278597156. Acesso em: 03 set. 2018. Há ainda uma versão legendada do vídeo destinada a pessoas com deficiência auditiva, disponível em: https://vimeo.com/278597666. Acesso em: 03 set. 2018. Outros dois textos foram utilizados para a elaboração do vídeo aqui referido; há traduções dos textos para o português, disponíveis em: http://cridi.ici.ufba.br/institucional/arquivos/artigos/artigo-sobre-o-processo-de-criacao-de-uma-Estrutura-para-aplicacao-do-OAIS-a-preservacao-digital-distribuida.pdf & http://cridi.ici.ufba.br/institucional/arquivos/artigos/artigo-sobre-o-uso-de-um-modelo-OO-IO-como-suporte-as-auditorias-e-analises-de-OAIS-colaborativos.pdf. Acesso em: 10 set. 2018.

[3] Online Computer Library Center Digital Archive. Cf. https://www.oclc.org/content/dam/support/digitalarchive/DigitalArchiveGettingStartedGuide.pdf. Acesso em: 05 set. 2018.

[4] Cf. http://www.ironmountain.com. Acesso em: 05 set. 2018.

[5] Kopal Long-Term Digital Information Archive. Cf. http://kopal.langzeitarchivierung.de/. Acesso em: 05 set. 2018.

[6] Lots of Copies Keep Stuff Safe. Cf. https://www.lockss.org/. Acesso em: 05 set. 2018.

auditar e acessar bits, enquanto as instituições participantes retêm responsabilidades relacionadas à lógica do conteúdo. Assim, as instituições continuarão encarregadas de toda a estruturação e empacotamento de dados e metadados, e da seleção de estratégias apropriadas de preservação. Isto é, por exemplo, modelado num sistema baseado no Fedora[7], na The State and University Library, o qual, no entanto, ainda precisa de preservação de bits para seus objetos.

À primeira vista, estabelecer requisitos para um repositório de bits compartilhado parece ser uma simples questão de estabelecer interfaces comuns nas operações de admissão-armazenamento e armazenamento-acesso. No entanto, uma análise dos requisitos das partes interessadas e envolvidas, revelou que um dos desafios de estabelecer um repositório de bits compartilhado é que as instituições têm requisitos substancialmente diferentes, quando se trata de preservação de bits. Os requisitos diferem de acordo com sua missão e com as coleções (fontes de dados) que elas guardam, assim como com a legislação nacional e internacional, tais como leis de direitos autorais e leis de proteção dos arquivos e dados pessoais. Exemplos são os requisitos para hardware específico para proteger dados ou diminuir custos, e para o acesso online rápido, enquanto outros podem estar armazenados em uma mídia offline. Coleções podem variar no que diz respeito ao seu valor percebido. Livros originalmente digitais podem, por exemplo, ter um valor mais alto que cópias digitais de livros impressos, porque estes últimos podem ser re-digitalizados, caso as cópias digitais sejam danificadas ou perdidas (RIEGER, 2008, p.42). Portanto, as instituições podem também estar dispostas a aceitar um nível de segurança menor para coleções de livros digitalizados. Estes níveis de serviços podem ser documentados dentro da norma PREMIS, a qual tem uma entidade denominada “nível de preservação” para descrever possíveis serviços de preservação, e o contexto no qual estes se aplicam (PREMIS, 2012, p.33-38). Há muitos desafios para atender a requisitos diferenciados em um sistema. Por exemplo: dados confidenciais, mesmo criptografados, podem ser armazenados junto com outros dados, tais como dados da *web*, que serão processados para estatisticas? Outro exemplo é como permitir e fazer exclusão segura ou sobrescrever dados de uma fonte, caso eles estejam

[7] Cf. http://fedora-commons.org. Acesso em: 05 set. 2018.

armazenados junto com dados de outra fonte, que precisam ficar inalterados sempre.

Este artigo fornece uma estrutura sistemática para analisar esses requisitos bastante complexos, usando o Modelo de Referência OAIS (CCSDS, 2012)[8]. Nós propomos o uso de uma visualização no estilo OAIS, para descrever as interfaces entre as instituições participantes e o repositório de bits. Esta análise destaca que não se pode igualar a funcionalidade de um repositório de bits com aquela da entidade funcional do OAIS, o Arquivamento. Ela mostra que são necessários subconjuntos de funções de todas as entidades funcionais OAIS, e é por isso que o denominamos de 'repositório compartilhado de bits' em vez de apenas uma instalação de armazenamento. O artigo apresenta, então, a arquitetura para um repositório nacional de bits da Dinamarca, que deve suportar múltiplas instituições com serviços diferenciados; por exemplo, integridade dos bits, confidencialidade, acessibilidade e processamento de dados. Isso inclui uma descrição dos desafios em especificar requisitos, como na segurança dos bits. E, por último, debatemos e finalizamos os resultados e apontamos para o trabalho futuro.

2 ANÁLISE DE REQUISITOS BASEADA NO MODELO OAIS

As instituições dinamarquesas se esforçam para estar em conformidade com o OAIS e, portanto, é um requisito importante que o repositório de bits também esteja em conformidade com o OAIS. O Modelo OAIS provou ser útil para discutir questões relacionadas com preservação de longo prazo e para analisar sistemas de repositórios. Como mostrado na Figura 1, o OAIS é estruturado em seis principais entidades funcionais – *Admissão, Arquivamento, Gerenciamento de Dados, Administração, Planejamento para Preservação e Acesso;* cada uma destas consiste numa série de funções.

[8] Ver tb. ISO 14.721 (2012).

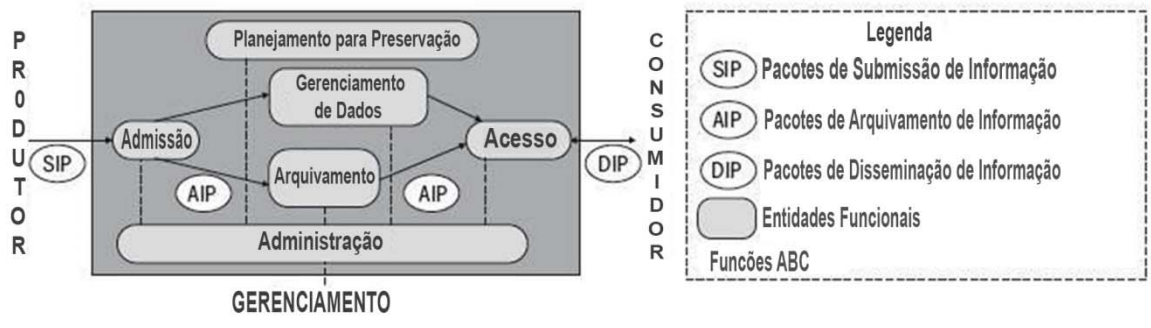

Figura 1: O Modelo de Referência OAIS

A Seção 6 da norma OAIS, 'Interoperabilidade dos Arquivos', discute a cooperação entre múltiplos arquivos OAIS. Inclui um exemplo de como as entidades funcionais Arquivamento e Gerenciamento de Dados podem ser compartilhadas, padronizando as interfaces de admissão-arquivamento e operações de acesso-arquivamento, onde o GERENCIAMENTO é encarregado dos acordos feitos entre os arquivos. Um design semelhante pareceria imediatamente um conceito apropriado para analisar requisitos para um repositório de bits compartilhado. No entanto, mais funções precisam entrar em ação, além das interfaces acima mencionadas e daquelas descritas em *Arquivamento*. O *Arquivamento* recebe solicitações de armazenamento e AIPs da Admissão, seleciona a mídia, prepara volumes e devolve a confirmação de armazenamento, incluindo um identificador. O *Arquivamento* também está em conformidade com níveis especiais de serviço ou medidas de segurança. No entanto, estas operações dentro do *Arquivamento* dependem de *input* de outras entidades funcionais: A Admissão pode indicar a frequência de utilização de dados necessária (tipo de mídia) por meio da solicitação de armazenamento. A Administração fornece politicas de gerenciamento de armazenamento e politicas de recuperação de desastres e, através do Planejamento para Preservação, também fornece recomendações sobre a evolução de sistemas e migração de mídia. Da mesma forma, o Gerenciamento de Dados é indiretamente encarregado de gerenciar os identificadores (*tokens*) e possivelmente trilhas de auditoria. Isso mostra que o Arquivamento está, na verdade, inter-relacionado com um subconjunto de funções de todas as entidades funcionais no OAIS.

Nós propomos, então, entender um repositório de bits como um subconjunto de um OAIS completo. Propomos visualizar o repositório de bits (BR, na sigla original em inglês) como incorporado na entidade funcional *Arquivamento* de cada

um dos repositórios institucionais (IR, na sigla original em inglês), como mostrado na Figura 2. Nós designamos esta visualização estilo OAIS de 'modelo IR-BR'. Mostraremos que o modelo IR-BR ajudará a analisar as interfaces e descrever a interoperabilidade entre múltiplos repositórios institucionais e o repositório de bits, e determinar a qual destas duas camadas pertencem os requisitos. É importante notar que o *Arquivamento*-IR não presume que um *AIP*-IR seja armazenado como um único objeto digital. Como mostrado na Figura 3, um *AIP*-IR produzido a partir de um *SIP*-IR pode consistir em mais objetos digitais, isto é *SIPs*-BR. Cada *SIP*-BR pode, por exemplo, ser uma representação de componentes de arquivos, metadados e etc. É responsabilidade do *Gerenciamento de Dados*-IR acompanhar as relações entre os *SIPs*-BR armazenados e suas informações de identificação, para um eventual acesso.

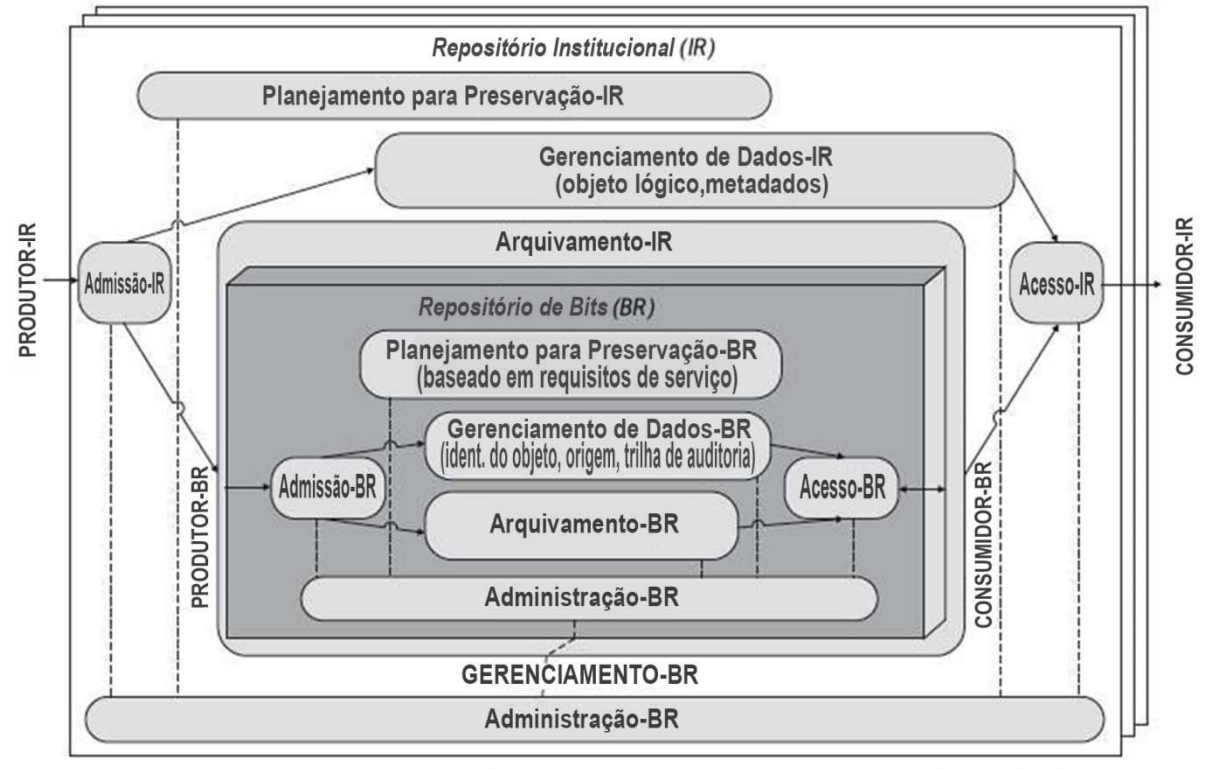

Figura 2: Mostra a visualização estilo OAIS, na qual o BR é compartilhado por múltiplos IRs, ao incorporar o BR no Arquivamento-IR.

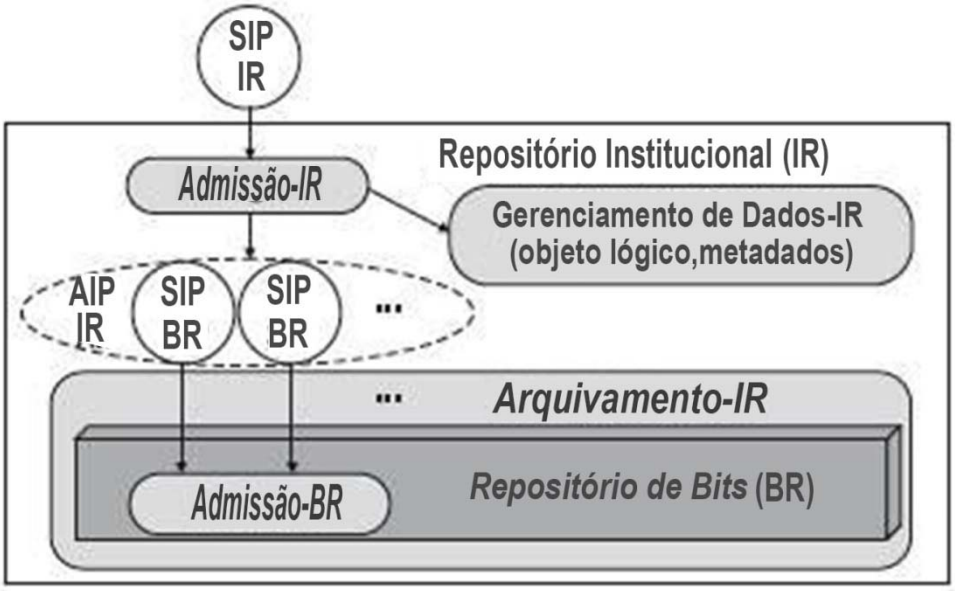

Figura 3: Exemplo de como um AIP-IR pode ser armazenado como várias SIPs-BR

Para analisar o modelo IR-BR em termos do OAIS, precisamos olhar mais cuidadosamente para as interfaces entre a *Admissão*-IR e *Admissão*-BR, e o *Acesso*-IR e o *Acesso*-BR, e como estas influenciam o modo como as entidades funcionais OAIS são decompostas em funções. A Figura 4 ilustra o fluxo entre essas interfaces.

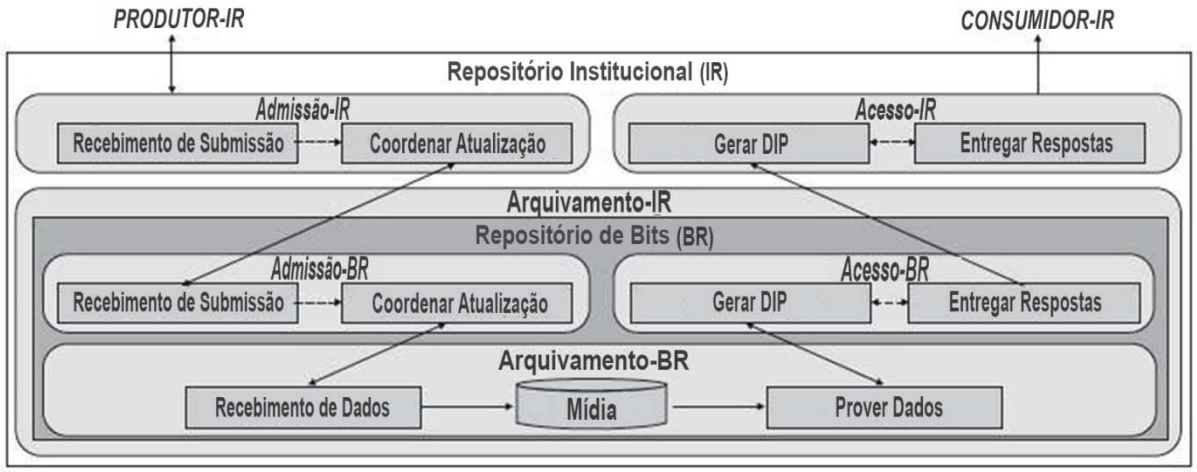

Figura 4: Mostra as interfaces dentro do Modelo IR-BR

As funções *Admissão*-IR são todas executadas dentro do IR. No entanto, quando um *AIP*-IR é transferido do *Arquivamento*-IR, isto é, admitido (*ingested*) no BR, ele faz um caminho ligeiramente diferente, nesta visualização estilo OAIS, do que se o BR fosse apenas *Arquivamento*. Neste, ele segue através das funções

Admissão-BR até chegar ao *Arquivamento*-BR. Da mesma forma, o recibo pelos dados aceitos e o armazenamento completo, é devolvido ao IR, enquanto o BR funciona como um *Arquivamento* de bits que está em conformidade com o OAIS. Assim, a *Admissão*-IR recebe a confirmação de armazenamento da *Admissão*-BR. Reparem que a função *Admissão*-BR/*Gerar AIP* é relativamente simples, já que não existe lógica anexada ao *SIP*-BR, exceto pela informação sobre identificação e pela possível documentação sobre trilhas de auditoria (isto é, nenhuma informação de representação é necessária, uma vez que olhamos apenas para fluxos de bits, não importando suas interpretações).

De forma similar, no *Acesso*-IR, todas as funções envolvidas são executadas dentro do IR, exceto, de novo, na interface entre o *Acesso*-IR e o *Acesso*-BR. Assim, a função *Acesso*-IR/*Gerar DIP* não resgata o *AIP*-IR diretamente, mas através da função *Acesso*-BR. Funções correspondentes são executadas no nível BR.

As funções de *Gerenciamento de Dados* são aplicadas dentro do IR e do BR respectivamente. Caso a instituição queira que as informações do *Gerenciamento de Dados*-BR sejam parte do *Gerenciamento de Dados*-IR, como por exemplo, informação sobre trilhas de auditoria, então estas informações devem passar pelo *GERENCIAMENTO*-BR.

Da mesma forma, as funções de *Administração* são gerenciadas dentro do IR e BR. No entanto, há uma exceção para a *Administração*-IR, especificamente: o recibo de relatórios sobre estatísticas operacionais da função *Arquivamento*-IR/*Gerenciar Hierarquia de Armazenamento,* via *Admissão*-IR *e Acesso-IR,* atravessa o BR. Além disso, como já descrito, os requisitos para o gerenciamento de armazenamento e políticas de recuperação de desastres, estabelecidos pela *Administração*-BR, são coordenados com a *Administração*-IR via *GERENCIAMENTO*.

Planejamento para Preservação está dividido entre o IR e o BR. O *Planejamento para Preservação*-IR, entre outras coisas, cobre a função *Monitoramento Tecnológico*, no que diz respeito à migração de formato de dados, enquanto migração de mídia é colocada na função *Planejamento para Preservação*-BR/*Monitoramento Tecnológico*. Reparem que isto também quer dizer que todas as atividades relacionadas com preservação funcional são responsabilidade unicamente do IR.

Examinando os aspectos pormenorizados de compartilhar um BR, a *Admissão*-BR deve ser capaz de receber *SIPs*-BR de IRs diferentes. Mais ainda, os IRs individuais devem ser capazes de consultar *SIPs*-BR como um *DIP*-BR, num estágio mais avançado, o que significa que os *SIPs*-BR devem, no mínimo, conter os bits a serem armazenados, uma identificação da origem dos dados e um identificador único dentro do IR. Reparem que estes requisitos podem também ser necessários para um único IR, se ele preservar mais fontes de dados. Como resultado, a função *Admissão*-BR/*Recebimento de Submissão* é solicitada a distinguir de quais fontes disjuntivas vêm os *SIPs*-BR, e os requisitos associados para armazenar o objeto digital.

De um modo geral, todos os requisitos que as instituições têm são baseados em orientações do *GERENCIAMENTO*-IR e resolvidos pela *Administração*-IR. Da perspectiva do BR, a *Administração*-IR representa o *GERENCIAMENTO*-BR. Assim, é na interface entre *Administração*-IR e a *Administração*-BR que o mapeamento dos requisitos para o BR acontece.

3 REQUISITOS PARA A ARQUITETURA DE UM REPOSITÓRIO DE BITS NACIONAL

Nesta seção nós descrevemos como o Modelo OAIS impactou a arquitetura do repositório nacional de bits compartilhado (DK-BR). No entanto, para entender o contexto, começaremos com uma breve descrição dos requisitos mais importantes para o DK-BR. A seguir, descrevemos a arquitetura geral e defendemos as escolhas feitas, baseadas em experiências. Visto neste contexto, descrevemos questões relevantes apontadas pela análise OAIS. Além disso, descreveremos os desafios restantes para expressar requisitos de segurança dos bits em geral, e determinar o nível correto de abstração.

3.1 REQUISITOS

A arquitetura do DK-BR é baseada numa análise dos requisitos das partes interessadas envolvidas, na análise OAIS, nas experiências dos próprios parceiros com a execução de sistemas de preservação de bits e pela observação de instituições internacionais similares.

Os repositórios de bits dinamarqueses existentes, como aquele usado para armazenamento na *web*, e também o DK-BR, todos eles se baseiam em considerações gerais para preservação ativa de bits, isto é, o numero de cópias, o grau de independência entre as cópias, e a frequência de auditorias de integridade dos bits (ROSENTHAL, 2010). A frequência de auditoria revela a frequência das somas de verificação de bytes (*checksum*) em andamento, que requer pelo menos três *checksums* para os dados armazenados, para que sejam comparadas e, caso se detecte alguma diferença, a cópia errada é então identificada através de uma votação, e deve ser substituída (CHRISTENSEN, 2005).

Da análise dos requisitos, ficou clara a necessidade de uma arquitetura bastante flexível para atender a requisitos diversos, com custos diferenciados. Por exemplo, havia requisitos para diferentes níveis de segurança dos bits, no qual o número de cópias tem influência direta nos custos. O mesmo se aplica para outros requisitos, como o acesso, onde havia requisitos muito caros, para ter uma cópia armazenada numa plataforma distribuída, a fim de permitir o processamento para estatísticas sobre dados da *web*, por exemplo, análise linguística para pesquisadores. Este processamento era exigido dentro do BR, porque a grande quantidade de dados tornava economicamente inviável e tecnicamente difícil de executar fora do BR. Outro exemplo é que o The Danish National Archives tem uma cópia dos dados offline em mídia ótica (DVD) devido a requisitos de acesso pouco frequentes e porque esta mídia é relativamente barata e útil para garantir a confidencialidade dos dados. Reparem também que os requisitos podem entrar em conflito; por exemplo, requisitos de confidencialidade e segurança dos bits. Enquanto mais cópias aumentam a segurança dos bits, o risco de comprometer a confidencialidade também aumenta.

3.2 UMA ARQUITETURA FLEXÍVEL PARA ATENDER VARIADOS REQUISITOS

A arquitetura para o DK-BR tem similaridades com sistemas como o LOCKSS e DuraCloud.

No entanto, o DK-BR difere, mas também se fortalece, na forma pela qual enfrenta requisitos diferenciados para segurança dos bits e de custos. Uma vantagem adicional é que evita ter um índice mestre central para o conteúdo BR, o

que minimiza o risco de ponto-único-de-falha. A arquitetura geral do DK-BR é ilustrada na Figura 5.

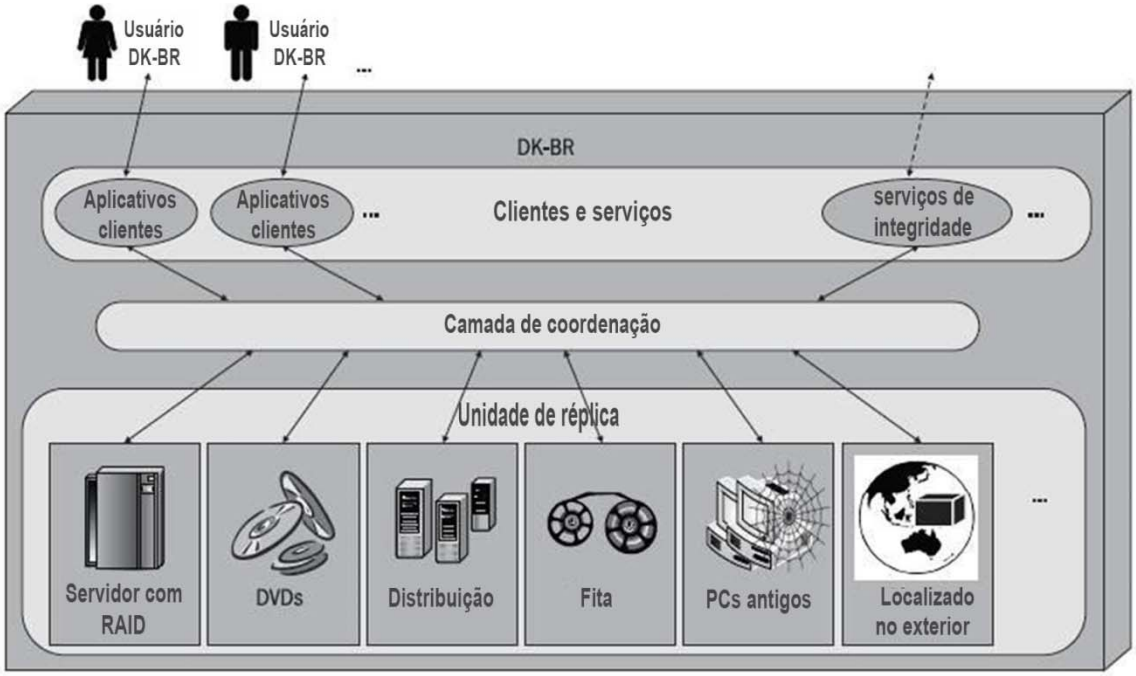

Figura 5: Arquitetura geral do DK-BR

As unidades de réplica formam o armazenamento básico para dados. Cada unidade de réplica consiste numa mídia e infraestrutura próprias, dando-lhe características específicas que são usadas quando um acordo de níveis de serviço (SLA, na sigla original em inglês) é definido para um usuário-BR. Uma unidade de réplica é definida como uma representação de uma cópia de dados que pode ser vista e analisada como uma unidade individual no nível abstrato. Para obter independência entre cópias, as unidades de réplica não têm conhecimento umas das outras. As unidades de réplica descritas são apenas exemplos ilustrados de unidades de réplica representando diferentes características, na forma da mídia e nas locações físicas e, assim, implicitamente, seus custos. Outras características podem ter relação com a mídia (por exemplo, taxa de transferência não magnética e erro de bit), organização, etc. As unidades de réplica podem, internamente, realizar ações de preservação, tais como migrações de mídia, desde que ainda satisfaçam as características especificadas. Os aplicativos clientes lidam com funções relacionadas ao usuário, como informações de admissão, acesso, gerenciamento e ações corretivas, enquanto os serviços podem, por exemplo, ser as auditorias de

integridade de bits (checagem da soma de verificação de bytes, ou *checksum checks*) e monitoramento. Aplicativos clientes e serviços são configurados para atender os SLAs (acordos de níveis de serviços) individuais. Mais informações podem ser encontradas em Jurik et al. (2012) ou em http://BitRepository.org.

Por fim, a camada de coordenação gerencia a comunicação entre aplicativos de usuários e unidades de réplica. Sua principal funcionalidade é ser o condutor para a comunicação.

A arquitetura é flexível para atender diferentes SLAs (acordos de níveis de serviços) para usuários, oferecendo diferentes combinações de unidades de réplica para melhor se adequar aos requisitos do cliente. A arquitetura suporta representações de diferentes unidades de réplica, que podem ser combinadas para atender requisitos especiais, por exemplo, para velocidade de acesso, ou processamento ou para atender requisitos de custos, ou para prevenir riscos específicos. Em outras palavras, a seleção de unidades de réplica para um SLA (acordo de níveis de serviços) é baseada numa análise das características da unidade de réplica individual e a combinação da capacidade das características da unidade de réplica de atenderem aos requisitos dos usuários. Por exemplo, a distância entre as unidades de réplica influenciará o risco de duas cópias serem destruídas ao mesmo tempo. A diversidade pode estar em muitos aspectos – tecnológico (hardware e software, mídia online e offline, mídia ótica e magnética, utilização de diferentes técnicas de armazenamento e utilização de diferentes vendedores), geográfico, organizacional, ou relacionado a dados (*checksums* baseados em diferentes algoritmos de *checksum*, características diferentes que expressam políticas de armazenamento somente leitura, ou nenhum processamento). Tais considerações foram utilizadas no repositório de bits existente para material dinamarquês da web em Netarkivet (CHRISTENSEN, 2005), mas não de maneira tão ampla quanto no DK-BR.

Outra diferença é que a unidade de réplica pode conter apenas a representação de uma cópia dos dados, na forma de um *checksum* derivado. Armazenar um *checksum* pode não fazer muito sentido por si, mas fornece um votante em uma solução de armazenamento com verificações de integridade de bits, mas com menos cópias, e assim, mais econômica. Numa análise de risco da parte de armazenamento do Netarkivet, economia foi um argumento para se ter apenas duas unidades de réplica com cópias completas e uma terceira com um *checksum*

derivado, a fim de obter três votantes. É importante, no entanto, considerar onde está armazenado o *checksum* derivado. Isto ficou evidente numa avaliação do Netarkivet usando o DRAMBORA, uma ferramenta para auditoria interna de repositórios digitais (DCC; DPE, 2007). Ele revelou que o *checksum* foi posicionado no mesmo local físico de uma das unidades de réplica, o que significava que um incêndio ou uma explosão resultariam na perda de dois votantes em três.

A arquitetura evita o ponto-único-de-falha, evitando um índice mestre central do conteúdo do BR. Em vez disso, ele é baseado em informações em cache que podem ser atualizadas ou verificadas por informações das unidades de réplica. Dessa maneira, não existe um único ponto que seja indispensável no cálculo do índice mestre. A arquitetura também evita o ponto único de falha ao ter instâncias paralelas duplicadas de camadas de coordenação e aplicativos de usuário, embora isto não seja mostrado na Figura 5. Também atende aos requisitos de escalabilidade e tempo de atividade, já que eles são projetados para serem substituíveis.

Outras ações de prevenção de risco são levadas em consideração, incluindo medidas de prevenção de riscos decorrentes das normas exigidas pelas instituições dinamarquesas, chamado DS 484, que corresponde a ISO/IEC 27.001 (2005). Isto envolve, por exemplo, acesso restrito aos servidores físicos, e separação dos ambientes de produção e teste.

3.3 VISUALIZAÇÃO OAIS

Numa visualização OAIS, todos os requisitos dos IRs são na prática, administrados pela *Administração*-BR, e traduzidos em requisitos específicos para as diferentes funções OAIS. Assim, requisitos específicos para o *Arquivamento*, tais como a seleção da mídia de armazenamento, taxa de transferência e taxa de erro de bit são tratados na *Administração*-BR, especificando um SLA (acordo de níveis de serviços) que define combinações apropriadas de unidades de réplica e suas características.

A *Admissão*-BR e o *Acesso*-BR operam apenas no processamento de dados na forma de sequências de bits, deixando estas funções próximas das funções ‘Recebimento de Dados’ e ‘Prover Dados’ do *Arquivamento* OAIS. No entanto, a *Admissão*-BR deve prover o *Gerenciamento de Dados*-BR com uma identificação exclusiva de sequências de bits e a origem à qual os fluxos de bits pertencem.

O aspecto mais complexo do *Gerenciamento de Dados* são as trilhas de auditoria. Na arquitetura DK-BR, trilhas de auditoria serão fornecidas, principalmente por unidades de réplica individuais, e isto deve ser considerado num desenho final – se, por exemplo, as informações sobre a migração de mídia devem aderir aos AIPs, ou fazer parte das informações de gerenciamento do sistema. Além disso, os requisitos sobre como as unidades de réplica devem preservar as trilhas de auditoria, devem ser declarados.

Como argumentado anteriormente, o *Planejamento para Preservação*-BR será realizado dentro do BR. Na arquitetura DK-BR, com diversas unidades de réplica independentes, isso realmente ocorre dentro de cada unidade de réplica, o que deixa o monitoramento tecnológico para mídia, para a unidade de réplica individual. Para garantir a segurança dos bits, as instituições têm que contar com documentação da *Administração*-BR, com relação a IDs, somas de verificação de bytes (checksums) e, possivelmente, trilhas de auditoria, isto é, será reportado via informações da *Administração*-BR à *Administração*-IR, e encaminhado ao *Planejamento para Preservação*-IR, onde ações podem ser iniciadas, em caso de alertas. Como o *Planejamento para Preservação*-IR não pode exigir diretamente que uma unidade de réplica, por exemplo, execute migração de mídia, o IR pode, em vez disso, mudar para outra unidade de réplica, dentro do BR, que melhor atenda aos seus requisitos. É também digno de nota que requisitos do Planejamento para Preservação-IR podem influenciar outros requisitos para o BR, como requisitos para migração de formato de arquivo. O motivo é que essa migração, em grandes volumes de dados pode exigir grande capacidade de processamento da CPU diretamente nos dados armazenados, uma vez que limites de capacidade, como largura de banda, pode tornar praticamente impossível fazê-lo fora do repositório, por meio do acesso e readmissão.

A organização do DK-BR influenciará a implementação da *Administração*-BR e as funções de relatório necessárias por meio ou em paralelo com os aplicativos clientes, mas ainda não está decidido como esta organização parecerá ao DK-BR.

3.4 ESPECIFICAÇÕES DE REQUISITOS ESPECIAIS E NÍVEL DE ESPECIFICAÇÃO

Requisitos para segurança de bits, confidencialidade, requisitos políticos e econômicos são difíceis de expressar de forma explícita sem envolver elementos dentro da arquitetura. Existem muitos desafios para especificar e controlar tais requisitos, o que significa que sempre haverá um elemento de confiança entre o IR e o BR.

Em geral, é desafiador expressar riscos por elementos quantitativos porque pode resultar em uma visão simplista. Os requisitos para segurança de bits são especialmente difíceis de expressar devido à falta de meios científicos para medir, verificar e controlar isto (ROSENTHAL, 2010). Como parte do trabalho com a arquitetura do DK-BR, nós tentamos projetar um modelo de risco. No nível da unidade de réplica, os riscos são indicados em termos de características da unidade de réplica, que são as mais explícitas possíveis. No nível BR, os riscos são indicados em termos de combinações e variações entre unidades de réplica para expressar a independência entre as cópias/unidades de réplicas. Além disso, o número de cópias e a frequência de auditoria de bits devem ser levados em consideração. O modelo de risco final poderia se basear em estatísticas ou em simulações, como feito para o Netarkivet (CHRISTENSEN, 2005). Adicionalmente, aspectos como ponto único de falha e aspectos da lista de verificação da norma Audit and Certification of Trustworthy Digital Repositories (ISO 16.363, 2012)[9] devem ser incorporados. Este trabalho não foi finalizado no pré-estudo, mas vários desafios relacionados foram identificados.

O nível de abstração especificando requisitos também pode ser um desafio. Quando a Netarkivet teve uma migração de hardware após cinco anos em produção, parecia que os requisitos originais eram muito específicos, incluindo a localização exata de uma unidade de réplica, e qual Sistema Operacional específico utilizar. A ideia original era exigir independência entre as unidades de réplica – na forma de distância e diferentes sistemas operacionais (OS) – mas isto não foi colocado como uma exigência, e não em termos de distâncias exatas entre as unidades de réplica e o quanto os dois Sistemas Operacionais devem ser diferentes. Consequentemente,

[9] NT: Ver tb. CCSDS, 2011.

um nível de abstração deve ser feito, o que evita dependências da evolução em curso, ou os requisitos devem ser ajustados regularmente para corresponder ao tempo presente.

4 DISCUSSÃO

Nossa análise do OAIS destaca que você não pode igualar a funcionalidade de um BR, com aquela da entidade funcional OAIS *Arquivamento*, e foi isto que levou ao modelo IR-BR. Se por um lado, não existe conflito na visualização do BR como um subconjunto de um arquivo OAIS, a visualização estilo OAIS, na qual o BR está incorporado ao *Arquivamento*-IR, pode estar sujeita a discussão. A razão é que algumas das funções que se comunicam através das interfaces entre o IR e o BR precisam ser redefinidas para levar em conta o fato de que o fluxo de dados e de informações tomam um caminho diferente. No entanto, a norma OAIS afirma que implementações reais podem agrupar ou desdobrar a funcionalidade de maneira diferente; assim, em nossa opinião, não comprometemos o conceito OAIS com essas redefinições.

O modelo IR-BR foi útil na identificação de interfaces e para entender o fluxo de dados e documentação entre o IR e o BR. Especialmente, o modelo ajudou a entender como gerenciar as trilhas de auditoria. Também mostrou que precisamos considerar a natureza dos identificadores dos objetos digitais. Prevendo ter um BR para os próximos 100 anos ou mais, pode não ser viável ter informações históricas relacionadas à origem, como informações relacionadas a uma instituição que já não mais existe, ocultas nos identificadores (ZIERAU; JOHANSEN, 2008). Portanto, deveria se considerar ter UUIDs (Identificador Universal Único) para objetos digitais no BR, junto com a origem dos dados e identificador fornecido pelo IR.

Se nós olharmos as outras variações de sistemas de preservação, como o LOCKSS ou o Kopal, a divisão entre o sistema deles e os clientes é colocada de maneira diferente, em comparação à apresentada neste artigo. O LOCKSS, por exemplo, também inclui alguma preservação funcional, enquanto nós analisamos um BR, sem qualquer conhecimento dos objetos digitais além das trilhas de identificação e auditoria. O Kopal também é descrito em termos do OAIS, mas deixa validação e empacotamento fora do repositório OAIS. A questão está em se o nosso modelo IR-BR pode ser usado também nestes casos. O objetivo do nosso trabalho

nunca foi tentar isto, mas nós esperamos que o mesmo modelo possa identificar locais de funcionalidade, interfaces e serviços, como fizemos para o BR neste artigo.

Embora o DK-BR tenha uma arquitetura flexível, as possibilidades finais de fazer SLAs (acordo de níveis de serviços) com usuários, dependerão do número e da variação de unidades de réplica implementadas dentro do DK-BR. Além disso, o modo como vários requisitos contradizem outros, especialmente observando a confidencialidade, integridade e custos, pode significar que há limites para as possibilidades.

Não importa o quanto os requisitos são bem expressos, o sucesso em separar preservação de bits e compartilhar ou terceirizar estas atividades, dependerá da capacidade do BR de demonstrar confiabilidade. Assim, a especificação de requisitos deve ser ampliada por iniciativas de auditoria e certificação, como a norma Audit and Certification of Trustworthy Digital Repositories (ISO 16.363, 2012)[10], que se estende à norma OAIS. Além disso, sem sanções ou procedimentos de escalonamento não é possível garantir que os requisitos sejam cumpridos.

5 CONCLUSÕES E TRABALHO FUTURO

Concluímos que é útil para as instituições que estejam configurando requisitos diferenciados para um repositório de bits compartilhado, visualizá-lo como um subconjunto de um OAIS completo. Neste sentido, o repositório de bits torna-se um repositório dentro dos repositórios das instituições, como descrito no modelo IR-BR. Esta análise ajudou a revelar locais de funcionalidades e configurar os requisitos. Além disso, descobrimos que os requisitos devem especificar como as trilhas de auditoria dentro do BR são gerenciadas e, possivelmente, enviadas para o IR. Descobrimos, ainda, que o BR deve fornecer informações de identificação e origem ao SIP-BR.

Para o DK-BR, descobrimos que existem diferenças significativas nos requisitos para diferentes fontes de dados. Para tornar economicamente vantajoso o compartilhamento de um BR, é importante buscar semelhanças para encontrar as

[10] NT: Ver tb. CCSDS, 2011.

áreas em que ganhos econômicos podem ser alcançados e definir a arquitetura que atenda aos requisitos diferenciados.

No trabalho com a arquitetura DK-BR, descobrimos que ainda é necessário aplicar um modelo de risco para expressar problemas, como segurança e confidencialidade dos bits, como requisitos explícitos. No entanto, sejam eles expressos através de um modelo de risco ou diretamente, o nível de abstração dos parâmetros ou requisitos deve ser feito com cuidado para garantir a preservação de bits e evitar complicar o caminho através de fatores como migração ou mudanças organizacionais ao longo do tempo. Tanto o modelo de risco quanto o nível de abstração para especificação de requisitos, podem ser assuntos para estudos futuros.

Ao expressar os requisitos para um BR, sempre haverá um elemento de confiança envolvido, mesmo que eles sejam expressos por meio de um modelo de risco ou outros meios. Portanto, o trabalho futuro incluirá a articulação de requisitos e especificações BR, em relação a certificação e iniciativas de auditoria. Isso também contribuirá para permitir o equilíbrio dos requisitos em relação aos custos.

AGRADECIMENTOS

Informações e discussões proveitosas vieram do envolvimento da The Royal Library no estudo de viabilidade dinamarquês para um repositório de bits nacional. Além disso, das discussões com Andreas Rauber, da Universidade de Tecnologia de Viena, *inputs* valiosos contribuíram para o artigo.

REFERÊNCIAS

CCSDS - CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS. **Audit and Certification of Trustworthy Digital Repositories**. CCSDS 652.0-M-1 Magenta Book. Washington, DC: CCSDS, 2011. Disponível em: https://public.ccsds.org/pubs/652x0m1.pdf. Acesso em: 06 set. 2018.

CCSDS - CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS. **Reference Model for an Open Archival Information System (OAIS)**. CCSDS 650.0-M-2 Magenta Book. Washington, DC: CCSDS, 2012. Disponível em: https://public.ccsds.org/pubs/650x0m2.pdf. Acesso em: 06 set. 2018.

CHRISTENSEN, N.H. Preserving the Bits of the Danish Internet. **5th International Web Archiving Workshop (IWAW05)**. VInena, Áustria, 22-23 set., 2005. Disponível em: http://netarkivet.dk/wp-content/uploads/iwaw05-christensen.pdf . Acesso em: 06 set. 2018.

DCC; DCP - DIGITAL CURATION CENTRE; DIGITAL PRESERVATION EUROPE. **Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)**. Version 1.0, [2008]. Disponível em: https://www.prestocentre.org/resources/tools/drambora-digital-repository-audit-method-based-risk-assessment. Acesso em: 06 set. 2018

ISO 14.721. **Space data and information transfer systems – Open archival information system (OAIS) Reference Model**. Genebra, Suíça: International Organization for Standardization, 2012.

ISO 16.363. **Space data and information transfer systems - Audit and certification of trustworthy digital repositories**. Genebra, Suíça: International Organization for Standardization, 2012.

ISO/IEC 27.001. **Information technology — Security techniques — Information security management systems — Requirements**. Genebra, Suíça: International Organization for Standardization, 2005.

JURIK, B.A.; NIELSEN, A.B.; ZIERAU, E.M.O. Flexible Bit Preservation on a National Basis. **Proceedings of the IS&T Archiving Conference 2012**. Copenhagen, Dinamarca, 12-15 jun., 2012, pp. 2-7.

PREMIS. Preservation Metadata Implementation Strategies. **Data Dictionary for Preservation Metadata**, Version 2.2, 2012. Disponível em: http://www.loc.gov/standards/premis/v2/premis-2-2.pdf. Acesso em: 06 set. 2018.

RIEGER, O.Y. **Preservation in the Age of Large-scale Digitization**: A White Paper. Council on Library and Information Resources (CLIR), CLIR Publication n. 141, 2008. Disponível em: https://www.clir.org/wp-content/uploads/sites/6/pub141.pdf. Acesso em: 06 set. 2018

ROSENTHAL, D.S.H. Bit Preservation: A Solved Problem? **International Journal of Digital Curation**, n. 1, v. 5, 2010, p.134-148. Disponível em: www.ijdc.net/article/download/151/224/. Acesso em: 06 set. 2018.

ZIERAU, E.; JOHANSEN, A.S. Archive Design Based on Planets Inspired Logical Object Model. In: HRISTENSEN-DALSGAARD, B.; CASTELLI, D.; JURIK, B.A.; LIPPINCOTT, J. (Eds). **Proceedings of the 12th European Conference - ECDL 2008.** Aarhus, Denmark, set. 2008, LNCS, v.5173, pp. 37–40. Heidelberg: Springer, 2008.

ZIERAU, E.; KEJSER, U.B. Cooperação interinstitucional num repositório de bits compartilhado. Traduzido para o português por Miguel Rio Branco Nabuco de Gouvea. Revisão de Rubens Ribeiro Gonçalves da Silva. Original em inglês publicado em **World Digital Libraries**: An International Journal, v.6, n.1, p. 25-36,

June 2013. Tradução para o português disponível em: http://cridi.ici.ufba.br/institucional/arquivos/artigos/artigo-sobre-o-compartilhamento-de-repositorios-de-bits-em-cooperacao-internacional.pdf. Acesso em: 10 set. 2018.

ZIERAU, E.; KEJSER, U.B. Cross-Institutional Cooperation on a Shared Bir-Repository. **World Digital Libraries**: An International Journal, v.6, n.1, p. 25-36, June, 2013. Disponível em: https://content.iospress.com/articles/world-digital-libraries-an-international-journal/wdl120098. Acesso em: 05 set. 2018.

ZIERAU, Eld; McGOVERN, Nancy Y. Supporting the Analysis and Audit of Collaborative OAIS's Using an Outer OAIS-Inner OAIS (OO-IO) Model. **iPRES 2014 – Proceeedings of the 11th International Conference on Digital Preservation**, Melbourne, 6-10 October, 2014, p.209-218. Disponível em: https://ipresconference.org/ipres14/sites/default/files/upload/iPres-Proceedings-final.pdf. Acesso em: 03 set. 2018.

ZIERAU, Eld; McGOVERN, Nancy Y. O Uso do Modelo OAIS Externo-OAIS Interno (OO-IO) para suporte à Análise e Auditoria de OAIS's Colaborativos. Original em inglês publicado em **iPRES 2014 – Proceeedings of the 11th International Conference on Digital Preservation**, Melbourne, 6-10 October, 2014, p.209-218. Tradução para o português disponível em: http://cridi.ici.ufba.br/institucional/arquivos/artigos/artigo-sobre-o-uso-de-um-modelo-OO-IO-como-suporte-as-auditorias-e-analises-de-OAIS-colaborativos.pdf . Acesso em: 05 set. 2018

ZIERAU, E.; SCHULTZ, M. Creating a Framework for Applying OAIS to Distributed Digital Preservation. **Proceedings of the 10th International Conference on Preservation of Digital Objects**, Lisboa, Portugal, 03-05 set., 2013, p.78-83. Disponível em: http://purl.pt/24107/1/iPres2013_PDF/iPres2013-Proceedings.pdf. Acesso em: 05 set. 2018.

ZIERAU, E.; SCHULTZ, M. Criando uma estrutura para aplicação do OAIS à preservação digital distribuída. Traduzido para o português por Miguel Rio Branco Nabuco de Gouvea. Revisão de Rubens Ribeiro Gonçalves da Silva. Original em inglês publicado em **Proceedings of the 10th International Conference on Preservation of Digital Objects**, Lisboa, Portugal, 03-05 set., 2013, p.78-83. Tradução para o português disponível em: http://cridi.ici.ufba.br/institucional/arquivos/artigos/artigo-sobre-o-processo-de-criacao-de-uma-Estrutura-para-aplicacao-do-OAIS-a-preservacao-digital-distribuida.pdf . Acesso em 05 set. 2018.