

Criando uma Estrutura para Aplicação do OAIS à Preservação Digital Distribuída¹

Eld Zierau
The Royal Library of Denmark
Søren Kierkegaards Plads 1
1219 København K
ph. +45 33 47 46 90
elzi@kb.dk

Matt Schultz
Educopia Institute
1230 Peachtree Street, Suite 1900
Atlanta, GA 30309
ph. +1 616 566 3204
matt.schultz@metaarchive.org

Resumo

O presente artigo descreve o trabalho em curso no sentido da criação de uma *Estrutura para Aplicação do Modelo de Referência para um Sistema Aberto de Arquivamento de Informação (OAIS) à Preservação Digital Distribuída (DDP)*. Essa Estrutura será útil nas futuras análises e(ou) auditorias de repositórios que realizam preservação digital de maneira altamente distribuída. Muito trabalho já foi realizado no sentido da Estrutura em si, incluindo revisões seletivas da comunidade de um “livro branco” (*white paper*), entrevistas com praticantes de DDP sobre estudos de casos e na análise do OAIS no que se relaciona à DDP. Este artigo vai fornecer informações básicas sobre esse conjunto de trabalhos, descrever a pesquisa realizada até a presente data, e descrever os componentes propostos para a Estrutura, incluindo conceitos e terminologia, posicionamento das entidades funcionais e os papéis e responsabilidades na realização de DDP.

Palavras-chave

Preservação Digital Distribuída; OAIS; Vocabulário; Entidades Funcionais; Papéis e Responsabilidades; Estrutura.

1 INTRODUÇÃO²

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 dos Autores. Cf. o original em 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: 06 set. 2018..

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-uso-de-um-modelo-OO-IO-como-suporte-as-auditorias-e-analises-de-OAIS-colaborativos.pdf> & <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

O presente artigo descreve o trabalho sendo realizado para criar uma *Estrutura para Aplicação do Modelo de Referência para um Sistema Aberto de Arquivamento de Informação (OAIS) à Preservação Digital Distribuída (DDP)*. Essa Estrutura será útil nas futuras análises e(ou) auditorias de repositórios que realizam preservação digital de maneira altamente distribuída. Na realidade, o uso da distribuição é muito comum no campo da preservação, mas não existia uma definição de “preservação digital distribuída” comumente aceita. Na medida em que o campo da preservação amadureceu, o termo “preservação digital distribuída” tem sido aplicado numa infinidade de abordagens de preservação, cobrindo iniciativas que constroem intencionalmente infraestruturas de distribuição como componentes essenciais de seus modelos de preservação, até outras iniciativas que apenas implementam medidas de backup para seus objetos digitais ou realizam todas as ações necessárias para preservação dentro da mesma organização. As razões para adotar abordagens distribuídas também podem variar no foco e na motivação (por exemplo, alcançar eficiência de custos através de recursos compartilhados (ROSENTHAL, 2010), suporte ampliado para grandes quantidades de dados (LITTMAN, 2006), e sustentabilidade generalizada do projeto, face a contingências e ameaças (ROSENTHAL, 2010).

Neste projeto de Estrutura, usamos o termo Preservação Digital Distribuída (DDP, na sigla original em inglês) para enfatizar a prática de aplicar distribuição de maneira intencional, tanto organizacionalmente quanto tecnicamente, para conseguir preservação digital, por exemplo, através de distribuição geográfica, heterogeneidade de infraestrutura, e diversidade organizacional. Uma definição mais concisa, para DDP é: o uso de *replicação, independência e coordenação* para abordar as ameaças conhecidas ao conteúdo digital, ao longo do tempo, para assegurar sua acessibilidade.

Em todo o mundo, numerosas iniciativas de preservação digital já estão utilizando metodologias, infraestruturas e aparatos organizacionais altamente distribuídos, para alcançar persistência confiável do conteúdo digital. Exemplos de tais iniciativas organizacionais e(ou) técnicas incluem Archivematica (<https://www.archivematica.org/>), the Danish Bit Repository (JURIK; NIELSEN; ZIERAU, 2012), Chronopolis (<https://library.ucsd.edu/chronopolis/index.html>), LuKII (www.lukii.hu-berlin.de/), LOCKSS (<http://www.lockss.org/>), UC3 Merritt (<https://merritt.cdlib.org/>), Data-PASS (<http://www.data-pass.org/>), DuraCloud

(<https://duraspace.org/duracloud/>), MetaArchive Cooperative (<https://metaarchive.org/>), DPN (<https://www.dpn.org/>), Internet Archive (<http://archive.org/>), iRODS (<http://www.irods.org>), e muitos outros.

Estas abordagens existentes, assim como aquelas que ainda estão para entrar em vigor, atualmente carecem de um vocabulário comum e estruturas conceituais para construir ambientes de preservação distribuída eficientes, confiáveis e auditáveis. Tal terminologia e modelos teóricos acordados conjuntamente ajudariam iniciativas para descrever e comparar suas infraestruturas e operações. Também ajudariam a aumentar a compreensão e conscientização do processo de DDP, tanto por praticantes como por organizações em busca de soluções para preservação. Finalmente, forneceria aos auditores (incluindo auto auditores) uma base fundamental para avaliar a confiabilidade de uma variedade de abordagens de distribuição.

O campo da preservação se apoia fortemente no Modelo de Referência para um Sistema Aberto de Arquivamento de Informação (OAIS, na sigla original em inglês) para fornecer modelos teóricos e abstratos e vocabulário para preservação digital (CCSDS, 2002)³. Essa norma OAIS fornece uma parte significativa da base para a Estrutura, que esta iniciativa pretende desenvolver. A Estrutura irá examinar mais detalhadamente o modelo OAIS para definir a metodologia e estrutura das ações tomadas em repositórios de preservação distribuída, organizacionalmente e(ou) tecnicamente. A Estrutura não pretende redefinir as normas existentes. Em vez disso, pretende apenas elaborar com modelos e vocabulários adicionais.

2 BASE DA ESTRUTURA

A consciência da necessidade desta Estrutura para Aplicação do OAIS à DDP surgiu de forma independente, tanto na América do Norte como na Europa. Na América do Norte, os debates iniciaram no começo de 2010, entre MetaArchive Cooperative, Chronopolis e a Biblioteca do Congresso. Mais tarde, em outubro de 2010, representantes da Biblioteca do Congresso e diversos grupos que adotam o LOCKSS se encontraram na *1st Annual Private LOCKSS Network (PLN) Conference*. Foi aqui que se formou um “grupo de notáveis” para debater as

³ Ver também ISO 14.721 (2003).

questões e lançar as bases do trabalho relacionado à Estrutura. Um Grupo de Trabalho formal foi constituído no início de 2011, para dar início à documentação de toda a gama de elaborações necessárias para ajudar a aplicar o OAIS ao ambiente DDP. Neste período o Grupo de Trabalho preparou uma Declaração de Intenções para a iniciativa, estabeleceu um espaço de trabalho para documentar casos de uso de DDP e análises de lacunas, apresentou um “livro branco”⁴, e solicitou a colaboração de diversos praticantes de DDP. Debates focados começaram no início de 2012 entre o Educopia Institute (<http://www.educoia.org/>) e a Biblioteca Real da Dinamarca (<http://www.kb.dk/>) para revisar o modelo desenvolvido pela Biblioteca, conhecido como o modelo IR-BR [9]. A Biblioteca Real da Dinamarca é pioneira na proposta deste modelo, que é uma abordagem com o objetivo de conseguir preservação distribuída confiável e auditável. O modelo IR-BR tem muitos conceitos e terminologias que se mostraram importantes para a Estrutura. Haverá mais sobre o modelo IR-BR ao longo deste artigo. O Grupo de Trabalho foi ampliado para incluir numerosas e bem conhecidas organizações, que compõem uma ampla variedade de casos de uso de DDP. Estas organizações incluem Archivematica, Chronopolis, Data-PASS, o Danish Bit Repository, DuraCloud, Internet Archive, LOCKSS, MetaArchive Cooperative, e UC3 Merritt.

2.1 A NECESSIDADE DE UMA ESTRUTURA

Em primeiro lugar, é importante reconhecer que o Modelo de Referência para um Sistema Aberto de Arquivamento de Informação (OAIS) não assume qualquer infraestrutura específica, técnica ou organizacional, mas antes prefere sintetizar as funções e requisitos de pacotes funcionais e de informação, que deveriam ser alcançados em qualquer implementação. Ele também descreve os papéis e responsabilidades assumidos por um arquivo, mas não especifica onde está essa responsabilidade, se num nível organizacional ou físico/geográfico.

Na prática, algumas instituições centralizaram as abordagens para construir um arquivo que esteja em conformidade com o OAIS – o que significa que, seus objetos digitais são admitidos e registrados em meios de armazenamento, localizados em uma locação, geográfica e organizacional, e administrada por um

⁴ NT: Um “livro branco” ou, no termo original, *white paper*, é um documento oficial publicado com a intenção de servir de informe ou guia sobre como enfrentar algum problema, por exemplo, técnico.

centro organizacional. Nesse paradigma, o local de interesse, responsabilidade e implementação é altamente centralizado.

Outras iniciativas usam abordagens distribuídas para construir um repositório ou rede – quer dizer que seus objetos digitais podem ser admitidos e registrados em meios de armazenamento, localizados em múltiplas locações geográficas, e que podem ser administrados através do uso de variados serviços de distribuição, oferecidos por diversas organizações, tudo para conseguir uma efetiva preservação de bits. Da mesma maneira, pode haver um forte argumento para a importância e relevância da distribuição no armazenamento, manutenção e aplicação de serviços adequados, como identificação de formatos, validação e guias de migração/padronização (por exemplo, registros de formatos como o PRONOM, UDFR, etc.). Como discutido mais abaixo (na seção 3.3.3, Terminologia), nesse paradigma a responsabilidade pode estar descentralizada e espalhada por várias locações físicas e(ou) institucionais. Embora o OAIS não seja de forma alguma contraditório com tais abordagens DDP, ele não define ou descreve explicitamente como os princípios e modelos OAIS se inserem nestas infraestruturas distribuídas – sejam elas organizacionais, geográficas ou baseadas em sistemas – cada uma das quais protege contra diferentes ameaças e que podem inclusive ser usadas em combinação. Uma estrutura comum pode ajudar a definir e dar sentido às coordenações adequadas.

Embora o OAIS comece a considerar questões de interoperabilidade entre diferentes arquivos (Seção 6, Interoperabilidade dos Arquivos), ele não aborda explicitamente o conjunto de interações que podem ocorrer entre entidades organizacionais separadas como parte do trabalho de um repositório de preservação digital distribuída, por exemplo, colaboradores que compartilham o gerenciamento de armazenamento por várias instituições diferentes (por exemplo, pelo uso de softwares embutidos peer-to-peer (P2P), microsserviços distribuídos, serviços geograficamente espalhados de armazenagem na Nuvem, ou outras configurações) ou até uma única organização que gerencia um arquivo composto de componentes de infraestrutura distribuídos – de novo, como no uso de microsserviços distribuídos da CDL & Archivematica.

Por esta razão, os primeiros praticantes de DDP se depararam com a necessidade de documentação adicional para descrever, com maior detalhamento, as diferentes funções, papéis e responsabilidades que emergem deste cenário de

distribuição. Tal Estrutura será útil em futuras análises e(ou) auditorias de repositórios que realizam preservação digital de maneira altamente distribuída.

2.2 MÉTODOS DE PESQUISA

Para abordar o desenvolvimento da Estrutura para Aplicação do OAIS à DDP, algumas atividades de pesquisa foram e continuam sendo executadas. Estas incluem; 1) a elaboração de um informe técnico foi divulgado para revisão pelas partes interessadas; 2) um conjunto completo de entrevistas sobre estudos de casos, com vários e diversificados praticantes de DDP; 3) uma análise detalhada do Modelo de Referência para um Sistema Aberto de Arquivamento de Informação (OAIS), visando preencher lacunas em conceitos e terminologia, posicionamento adequado das entidades funcionais e os papéis e responsabilidades para DDP; 4) uma revisão da literatura relacionada com DDP e OAIS.

2.2.1 Livro Branco (*White Paper*)

O desenvolvimento de um livro branco (*white paper*) tinha a intenção de demonstrar claramente a necessidade de uma Estrutura para Aplicação do OAIS à DDP. Ele abrangia muito do que foi abordado na seção 2.1 deste artigo. Ele mostrou ser um recurso de extrema utilidade para dar foco ao trabalho proposto, de construir a Estrutura em si, e para disseminar informação sobre a Estrutura entre os numerosos grupos interessados na DDP e especialistas em preservação digital.⁵ O *feedback* apurado com as revisões do informe técnico já desempenhou um papel fundamental no aperfeiçoamento do conteúdo e desenho da Estrutura. (ver seção 3, abaixo).

2.2.2 Entrevistas sobre Estudo de Casos

Também foram realizadas entrevistas sobre estudos de casos, com cada parceiro do Grupo de Trabalho, onde cada um respondeu uma série de questões consistentes, a saber:

⁵ Para solicitar uma cópia do livro branco (*white paper*) para a Estrutura, favor entrar em contato com Matt Schultz ou Eld Zierau (e-mails acima).

- Quais elementos da sua organização são distribuídos? (por exemplo, gerenciamento, infraestrutura de armazenamento, serviços de preservação).
- Qual o maior desafio em trabalhar de maneira distribuída?
- Se você usou qualquer ferramenta de auditoria (OAIS, TRAC⁶, DRAMBORA⁷, etc.), percebeu alguma lacuna entre conceitos e terminologia, nestas ferramentas e as formas como você realiza distribuição?
- Que deficiências, se houver, você percebe nestas ferramentas de auditoria, da forma como foram aplicadas ao seu ambiente distribuído?

As respostas a essas perguntas foram gravadas e usadas para identificar qualidades técnicas e organizacionais específicas e características de DDP que possam ser destacadas para a Estrutura. Esses elementos são discutidos na seção 3, abaixo.

2.2.3 Análise do OAIS

Documentar de forma eficiente, dentro da Estrutura, as partes relevantes do OAIS que tenham relação com DDP (e vice-versa), depende não só dos nossos estudos de casos, mas também através de uma análise minuciosa do OAIS em si. No núcleo do OAIS está um conjunto de conceitos de preservação digital e terminologia, entidades funcionais, e papéis e responsabilidades. O Grupo de Trabalho está revisando cada um destes elementos fundamentais do OAIS e procurando tanto associações quanto lacunas no que diz respeito à DDP, na sua melhor definição, tanto teoricamente como pelo estudo de casos existentes.

2.2.4 Revisão da Literatura

Muito trabalho de documentação foi realizado para descrever a aplicação correta do OAIS aos fluxos de trabalho de preservação digital e repositórios. De maneira similar, muito foi escrito sobre as diversas implementações distribuídas de preservação digital. Será vital que se incorporem informações relevantes de todas essas publicações existentes, para garantir o contexto adequado, continuidade e

⁶ Ver CRL – OCLC, 2007. Ver também ISO 16.363 (2012).

⁷ Ver DCC; DCP, 2008.

inteligibilidade da Estrutura para seu público alvo. Usuários da Estrutura irão se beneficiar com sua associação a este conjunto ampliado de informação.

3 ESTRUTURA – AGORA E NO FUTURO

Até o momento foram identificados alguns tópicos necessários para criar uma Estrutura que possa ser útil em futuras análises e(ou) auditorias de repositórios que realizam preservação digital distribuída. Estes tópicos incluem *terminologia* para DDP e apresentam uma série de *conceitos do mais alto nível, princípios e diretrizes*. Abaixo está a descrição do público alvo pretendido para a Estrutura, sua abrangência e possibilidades de publicação, assim como um resumo dos principais componentes propostos para a Estrutura, do conteúdo existente e o pretendido.

3.1 PÚBLICO-ALVO PRETENDIDO

Existem múltiplos públicos-alvo para essa Estrutura aos quais precisaremos prestar muita atenção. Eles abrangem uma série de instituições interessadas, responsáveis por e preocupados com a persistência da informação digital, assim como suas comunidades designadas, incluindo agências governamentais, bibliotecas públicas e arquivos e curadores de dados de pesquisa, entre outros.

O público-alvo principal – abrangendo cada uma destas partes interessadas – são aquelas organizações que procuram conjuntamente desenvolver ou aprimorar sistemas de preservação digital distribuída (DDP) e que necessitam de orientação sobre a maneira responsável de fazê-lo. Um segundo e igualmente importante grupo consiste nos órgãos de auditoria que procuram avaliar esses sistemas de DDP, e que se beneficiariam com as elaborações e interpretações fornecidas pela Estrutura. Por fim, existem aquelas organizações procurando depositar seus objetos digitais em sistemas assim, e procurando entender seu funcionamento. Pode também haver outros públicos-alvo, incluindo aqueles querendo acessar e usar objetos digitais.

3.2 ABRANGÊNCIA

A Estrutura atualmente procura abordar principalmente as diversas áreas relacionadas abaixo, com respeito tanto à DDP quanto ao OAIS. Como indicado

abaixo, na seção 4, Debate e Trabalho Adicional, haverá uma atenção especial, durante as fases posteriores de elaboração e revisão, sobre como melhor abordar a relação da distribuição com serviços funcionais de preservação, além daqueles em uma implementação mais geral de um repositório, a qual continua sendo o foco principal da Estrutura. Esses serviços funcionais de preservação fazem parte de um esforço geral de preservação digital, e abrangem coisas como serviços de registro de formatos e como estes podem ou não ser gerenciados e hospedados por múltiplas organizações (por exemplo, PRONOM, UDFR, etc.). O que provavelmente estará mais ou menos fora do escopo (pelo menos no início) é a incorporação de lições da comunidade geral de software de código aberto, onde colaboração, sustentabilidade e expansibilidade (em contrapartida a hospedar e manter tecnologias distribuídas e recursos administrativos) são mais importantes. A comunidade de código aberto continua sendo importante e essas lições provavelmente contêm muitos insights valiosos. Embora estejam fora do foco imediato e um pouco amplas para os efeitos da Estrutura atual, haverá um esforço para estudar como estas comunidades podem contribuir com a Estrutura final.

3.3 PUBLICAÇÃO

Essa Estrutura pode, potencialmente, tomar várias formas. Um exemplo que já existe é o *Producer-Archive Interface Model Abstract Standard (PAIMAS)*⁸, que é uma norma adicional do próprio OAIS. A adoção da abordagem desse documento necessitaria da revisão e aprovação pelo Comitê Consultivo para Sistemas de Dados Espaciais (CCSDS) e a International Organization for Standardization (ISO). Da forma como está sendo proposto, esse pode não ser o status mais apropriado para o trabalho. A Estrutura poderia, talvez mais apropriadamente, existir como um simples documento ou publicação, revisado pela comunidade, e hospedado e tornado acessível por uma organização respeitada. Além dessas formas tradicionais de documentos, a Estrutura pode existir como um recurso modular *web*. Pode haver outras formas. O Grupo de Trabalho continuará debatendo as formas adequadas de publicação, à medida que a elaboração da Estrutura prossegue. A elaboração da Estrutura está programada para ser realizada pelo Educopia Institute e pela

⁸ Ver também ISO 20.652 (2006).

Biblioteca Real da Dinamarca (financiada pelo Ministério da Cultura da Dinamarca), em conjunto com o Grupo de Trabalho durante o restante de 2013, com revisões pela comunidade programadas para o Outono/Inverno de 2013.

3.4 DESENHO

O desenho da Estrutura foi desenvolvido com base em conclusões de estudos de casos e revisões, sugestões do Grupo de Trabalho, revisões pelas partes interessadas do livro branco (*white paper*), assim como análises preliminares daqueles componentes do OAIIS mais relevantes para DDP (e vice-versa). Baseado neste desenho, a Estrutura propõe cobrir os elementos a seguir:

3.4.1 Introdução

Esta seção irá indicar de forma concisa o objetivo e os fundamentos para a Estrutura e fornecerá uma visão geral dos componentes.

3.4.2 Contexto e Visão Geral da DDP

A Estrutura propõe, em primeiro lugar, apresentar um conjunto completo de informações com contexto e visão geral sobre preservação digital distribuída (DDP), nos seus próprios termos, uma vez que ela evoluiu para se diferenciar das abordagens mais centralizadas de preservação digital, onde objetos digitais são admitidos e registrados em dispositivos de armazenamento localizados em uma única localização geográfica e organizacional e administrada por um único centro organizacional. Isso ajudará a orientar os leitores para esse ambiente específico e determinar os limites de uma aplicação efetiva do OAIIS, para os propósitos da Estrutura.

3.4.3 Terminologia

Muitos dos estudos de casos reforçaram a importância de ter uma terminologia aperfeiçoada para descrever as partes de um sistema distribuído. Alguns analistas expressaram apoio para a necessidade de uma terminologia

específica para a DDP, como alguns dos termos propostos e descritos abaixo. Como mencionado na introdução, esta é mais uma tarefa de expandir a terminologia do OAIS, do que, normativamente, redefinir qualquer terminologia OAIS existente. Mesmo assim, quando parecer adequado e útil, não se deixará de encorajar novas e melhores normatizações da terminologia para a comunidade de preservação digital.

Até o momento, surgiu a necessidade de definir dois tipos de termos:

- *Termos Principais*, como a definição de Preservação Digital Distribuída, que, na sua forma mais simples, é a que foi dada na introdução, baseada no uso de *replicação, independência e coordenação* (definições ampliadas também serão incluídas na Estrutura final)⁹;
- *Termos de Apoio* incluem o significado de *independência e coordenação*, e também termos como *unidade de armazenamento, nó de armazenamento, ambiente de armazenamento, cache/pillar*, assim como definições mais amplas de termos OAIS como *replicação e recuperação de desastres*.

A definição de DDP deriva em parte da experiência amadurecida de realizar preservação de bits da maneira correta, onde a segurança dos bits é melhor garantida através da correta coordenação de replicações independentes de dados. A independência das replicações requer que as réplicas sejam distribuídas geográfica e organizacionalmente, e coordenadas por rápida e efetiva verificação de integridade. A definição de DDP também deriva em parte da experiência amadurecida de realizar preservação funcional quando preservação sustentável depende de conhecimento compartilhado e soluções sobre como e quando fazer migração de formato. Aqui, a responsabilidade de registros de formato e o desenvolvimento de serviços confiáveis (por exemplo, microsserviço de migração) podem depender mais de seus hospedeiros nas diferentes organizações. Para saber como esta área de preservação funcional distribuída será explorada no desenvolvimento da Estrutura, ver a seção 4, Debate e Trabalho Adicional.

Exemplos típicos de termos de apoio à DDP são quaisquer um daqueles usados para a *unidade de armazenamento, nó de armazenamento, ambiente de armazenamento, cache e(ou) pillar*. Essa é a unidade que forma o armazenamento

⁹ Vamos seguir uma abordagem similar ao ALCTS (isto é, definições curtas, médias e longas). Veja aqui: <http://www.ala.org/alcts/resources/preserv/defdigpres0408>.

básico para uma cópia de dados num repositório de bits. Essa unidade é baseada em tecnologia específica com uma organização em torno dela, a qual pode ser responsável pelas operações básicas, monitoramento tecnológico, etc. A dificuldade, aqui, está em atingir um uso mais padronizado, já que a maioria dos termos têm outros significados também, e assim, podem ser mal interpretados, em contextos diferentes.

Replicação e Recuperação de Desastres são exemplos de termos de apoio que são extensões de termos OAIS. No OAIS, o termo Replicação primeiramente tem um contexto e significado de migração, enquanto no ambiente DDP é mais provável que haja um contexto e significado com medidas de integridade proativas, que seja mais próximo do OAIS, na forma como aborda Recuperação de Desastres. No entanto, recuperação de desastres como contexto e significado primário para replicação, também não faz completamente justiça ao seu uso em um ambiente DDP. A Estrutura deixará isso mais claro.

3.4.4 Relações DDP e OAIS

A Estrutura pode agregar valor para a DDP e para o OAIS, pelo seu foco no uso responsável da distribuição, nos níveis técnico e organizacional. As relações entre DDP e OAIS serão associadas uma à outra na Estrutura, ao se fazer uma ponte entre a perspectiva DDP e a perspectiva OAIS, especialmente através da análise do correto posicionamento das entidades funcionais, papéis e responsabilidades.

- *Entidades Funcionais* descreverão cenários para o posicionamento dos componentes de uma entidade funcional OAIS entre ambientes distribuídos, incluindo aqueles que implicam na coordenação/colaboração entre múltiplas organizações. Um exemplo pode ser encontrado no modelo IR-BR (ver *Modelos* abaixo).
- *Papéis e Responsabilidades* descreverão os papéis e responsabilidades em instituições que estão replicando e preservando informação digital, de maneira distribuída, geograficamente e(ou) organizacionalmente.

3.4.5 Estudos de Casos

O objetivo desta seção da Estrutura será apresentar alguns exemplos de estudos de casos, sobre como várias ações são modeladas e implementadas, de maneira distribuída, tanto organizacionalmente quanto tecnicamente. Não se pretende de forma alguma que estes estudos de casos sejam totalmente representativos do campo da DDP em geral, e nem se tem por objetivo a exclusividade mútua (isto é, alguns grupos DDP podem compartilhar as qualidades de estudos de casos com outros). Mesmo assim eles documentarão uma impressionante variedade de configurações que pode nos dar uma visão sobre as várias qualidades e características da DDP.

Os estudos de casos são baseados em entrevistas focadas no sentido de encontrar os elementos da DDP e OAIS, da forma como são utilizados pelas várias organizações e iniciativas técnicas resumidas abaixo.

Estudo de Caso Archivemática

Apoio à DDP baseado na comunidade

A colaboração técnica muitas vezes é essencial para apoiar uma implementação DDP. O código aberto e a infraestrutura de microsserviços baseados no OAIS da Archivemática é altamente dependente de uma comunidade distribuída e colaborativa, tanto de usuários quanto de desenvolvedores. O estudo desse caso destacará boas práticas para realizar abordagens técnicas coordenadas, no sentido de alcançar preservação digital, através dessas plataformas modulares e flexíveis.

Estudo de Caso Chronopolis

Equilibrando parcerias para a DDP

As instituições que compõe o programa Chronopolis estruturaram o fluxo de trabalho administrativo necessário para otimizar a distribuição de recursos entre três organizações muito heterogêneas. Esse estudo de caso descreverá a importância de posicionar e coordenar responsabilidades administrativas de forma efetiva, entre múltiplas organizações independentes, colaborando no sentido de uma infraestrutura de repositório distribuída e compartilhada.

Estudo de Caso Danish Bit Repository

Preservação colaborativa flexível de bits entre instituições para DDP

Esse estudo de caso discute uma plataforma para preservação compartilhada de bits, que permite distribuição de cópias de dados (não importando se a mídia está online ou offline) assim como os serviços neles realizados. Preservação compartilhada de bits é um desafio especial para os papéis e responsabilidades, em conexão com a forma pela qual a independência entre cópias de dados é mantida, e também sobre como garantir que sejam oferecidas soluções de preservação de bits para diferentes requerimentos de integridade, confidencialidade e disponibilidade de bits.

Estudo de Caso Data-PASS

Coordenando os 'stakeholders' para DDP

As principais partes interessadas envolvidas (*stakeholders*) na parceria Data-PASS conseguiram equilibrar independência institucional com compartilhamento e coordenação de recursos entre as instituições participantes da preservação colaborativa. Esse estudo de caso mostrará a importância do envolvimento das partes interessadas nos processos de decisão, de administrar as relações entre os parceiros, coordenar as operações, desenvolver práticas colaborativas, e participação numa infraestrutura comum, para coordenar de maneira efetiva as ações e garantir sustentabilidade do esforço geral organizacional de DDP.

Estudo de Caso DuraCloud

Alavancando a infraestrutura de nuvem para DDP

Esse estudo de caso discutirá a abordagem inovadora de gerenciar fluxos de trabalho vindos da admissão e armazenamento entre vários provedores de armazenamento em Nuvem, públicos e privados. O estudo de caso também discutirá o papel e a importância de auditar conteúdo externo e fazer uso efetivo dos acordos de serviço, no cenário de serviços da Nuvem.

Estudo de Caso Internet Archive

Papéis e responsabilidades específicas para DDP

Esse estudo de caso mostrará o esforço impressionante do Internet Archive em procurar parceiros internacionais, que possam preencher funções específicas no

conjunto geral de responsabilidades técnicas e organizacionais para realizar DDP para o Internet Archive. Muitos destes posicionamentos chave tiveram o objetivo de aperfeiçoar os processos necessários para gerenciar e replicar as grandes quantidades de dados do Internet Archive

Estudo de Caso MetaArchive

Formando uma Comunidade para DDP

Reunir múltiplas organizações para realizar preservação digital distribuída é uma oportunidade de compartilhar aprendizados e desenvolver tecnologias mutuamente benéficas. Criar tal comunidade de práxis entre múltiplas organizações pode ser tanto desafiador quanto gratificante. Esse estudo de caso da Cooperativa MetaArchive irá documentar as estruturas e abordagens que tornaram possível a formação de comunidades para este grupo DDP.

Estudo de caso UB3 Merritt

Serviços dedicados para DDP

Este estudo de caso mostrará como a Biblioteca Digital da Califórnia distribuiu seus serviços de repositório UC3 Merritt para otimizar o desempenho de sua arquitetura e fluxo de trabalho, orientados para microsserviços.

3.4.6 Modelos

A Estrutura também vai investigar diferentes modelos que podem ajudar a entender DDP. Um modelo que está sendo explorado pode ser visto na Figura 1, que ilustra o exemplo de distribuição para a preservação de bits (ZIERAU, 2011).

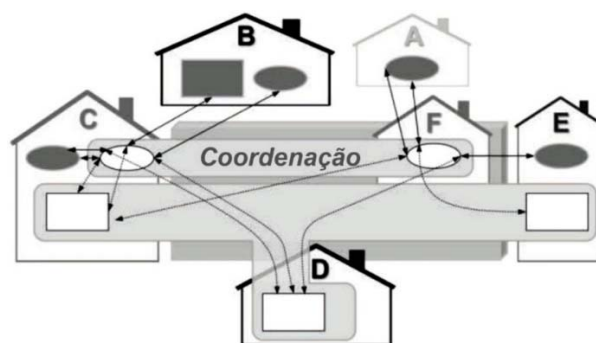


Figura 1: Distribuição para Preservação Digital

Os quadrados podem representar uma “unidade de armazenamento”, cada uma guardando uma cópia dos dados, as casas podem representar uma organização, e os círculos representam serviços e processos. Os círculos/quadrados escuros são totalmente internos, enquanto os brancos são parte de uma solução compartilhada distribuída. Neste caso, a “casa” A é apenas um consumidor de uma solução distribuída, que não tem nenhuma cópia dos dados em si. Por outro lado, a “casa” D tem apenas a função de provedor da “unidade de armazenamento” para guardar uma cópia dos dados. De maneira similar, a “casa” F é só um provedor de um serviço coordenado de processamento de serviços, dentro de uma solução distribuída. Isto é apenas para mostrar que DDP pode existir em vários formatos e com variados graus de complexidade.

Outro exemplo é o modelo IR-BR mostrado na Figura 2 (ZIERAU; KEJSER, 2010)¹⁰. Esse modelo foi desenvolvido em conexão com um estudo de viabilidade para a plataforma do repositório de Bits Dinamarquês (Danish Bit repository).

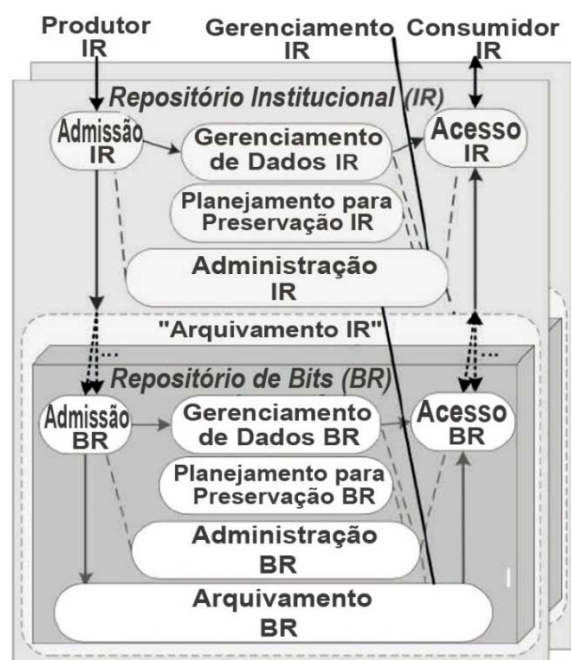


Figura 2: O modelo IR-BR

¹⁰ NT: Sugerimos acessar versão mais atual do referido artigo, publicada em 2013, com o mesmo título, disponível em: <https://content.iospress.com/articles/world-digital-libraries-an-international-journal/wdl120098>. Acesso em: 10 set. 2018. Este próprio artigo que ora se lê é a tradução para o português da versão mais atual em inglês.

O modelo IR-BR mostra um Repositório de Bits (BR, na sigla original em inglês) compartilhado como sendo um OAIS compartilhado por vários Repositórios Institucionais (IR, na sigla original em inglês), que também são repositórios OAIS. Portanto o BR é um OAIS dentro de muitos OAIS's.

A ideia é que cada IR tenha seu próprio Monitoramento Tecnológico como parte do seu *Planejamento para Preservação*, que cobre formatos de arquivos admitidos. No entanto, a totalidade do *Planejamento para Preservação* para o IR e seu uso do BR, também cobrirá Monitoramento Tecnológico como parte do *Planejamento para Preservação BR* e, portanto, cobrindo coisas como mídia de armazenamento. Além disso, o BR se apoia em diferentes “unidades de armazenamento” como mostrado na Figura 1. Quer dizer que o *Planejamento para Preservação* para um IR cobre seu próprio *Planejamento para Preservação IR* interno, assim como o *Planejamento para Preservação BR*, o qual pode ser distribuído por muitas organizações, com seus próprios monitoramentos tecnológicos das mídias usadas em suas “unidades de armazenamento”.

De maneira similar, novos modelos podem ser adicionados à Estrutura, caso eles agreguem valor nas análises e(ou) auditorias de DDP.

3.4.7. Aplicando Metodologias de Auditoria

Muitos dos estudos de casos abordam os desafios de se aplicar metodologias de auditoria aos seus OAIS distribuídos. Tanto o DRAMBORA quanto o TRAC se inspiram no OAIS, portanto faz sentido presumir que as extensões e descrições da aplicação de DDP ao OAIS podem ser usadas na abordagem de alguns destes desafios, que incluirão:

- *Tendências e Abordagens em Auditoria*: oferecerá uma visão geral das duas abordagens dominantes em auditoria (de cima para baixo com TRAC e de baixo para cima com o DRAMBORA) e colocará em destaque algumas das tendências e direções com a padronização das métricas de auditoria TRAC e o surgimento de agências de auditoria;
- *Métricas TRAC – Como se aplicam?*: abordará a maneira como as áreas do TRAC são melhor interpretadas para ambientes DDP. Esta descrição também

abordará qualquer área potencial específica de DDP que o TRAC atualmente não aborda plenamente;

- *Audidores TRAC – Como Abordar DDP?*: Tratará da forma como auditores podem se preparar, para aplicar métricas TRAC num ambiente DDP;
- *DRAMBORA – Como Abordar DDP?* Sobre como o DRAMBORA pode ser melhor interpretado e usado em ambientes DDP.

3.4.8. Conclusão – Usando a Estrutura

Esta seção irá resumir e fornecer um conjunto de diretrizes e sugestões para aplicar a Estrutura, tanto para os atuais praticantes de DDP como para aqueles interessados em usar essas abordagens.

4 DEBATE E TRABALHO ADICIONAL

Embora tenham sido realizadas muitas pesquisas, e a Estrutura proposta pareça bem definida, baseada em contribuições do Grupo de Trabalho e feedback da comunidade mais ampla de partes interessadas envolvidas (*stakeholders*), ainda existem questões relevantes e tópicos a serem explorados, à medida que prossegue o esforço para a elaboração e as etapas adicionais de revisão. A iniciativa Estrutura pretende continuar a expandir o debate sobre esse tópico e também sobre a participação dos *stakeholders*, através de posteriores publicações de conferências, sessões de pôsteres e eventos organizados.

Entre as questões e tópicos necessitando de atenção adicional, como mencionado acima, estão a(s) forma(s) final(ais) que a Estrutura deveria assumir, e onde deveria se hospedar para uma mais ampla cobertura e manutenção ao longo do tempo. Como também foi mencionado acima, nós reconhecemos que análises de distribuição para preservação digital, no contexto de um sistema e ambiente de repositórios, deveriam ser ligeiramente diferenciadas de coisas como colaboração para registro de formatos e outros desenvolvimentos de tecnologias de código aberto. Definir os limites adequados também é uma questão que provavelmente merece mais atenção e debate.

Mais trabalho também será necessário para compreender a função e as necessidades dos consumidores de objetos digitais e requisitos confiáveis para acesso em um ambiente de preservação digital distribuída.

Finalmente, já que esta iniciativa Estrutura não tem como objetivo ocupar o lugar de uma norma, os autores da Estrutura estão cientes de que a devida atenção será dada para tornar essas relações claras em todo o processo.

5 CONCLUSÃO

Este artigo expôs os primórdios de uma Estrutura para *Aplicação do Modelo de Referência para um Sistema Aberto de Arquivamento de Informação (OAIS, na sigla em inglês) à Preservação Digital Distribuída (DDP, em inglês)*.

Esta é uma iniciativa internacional, composta por numerosos praticantes da DDP e partes interessadas envolvidas (*stakeholders*) que vêm trabalhando contínua e colaborativamente, desde 2011, para definir claramente as áreas que precisam ser trabalhadas, as quais incluem pesquisa e documentação dos mapeamentos corretos entre OAIS e DDP nas áreas de *terminologia, entidades funcionais, e papéis e responsabilidades*.

Este trabalho resultou num 'livro branco' (*white paper*), que passou por revisão preliminar por vários especialistas em DDP e preservação digital; numa série de entrevistas, em estudos de casos, com praticantes de DDP; numa análise do OAIS com relação à DDP; e numa revisão minuciosa da literatura.

A Estrutura propõe incluir vários elementos úteis, como Contexto e Visão Geral da DDP, debates sobre Terminologia, Relações entre DDP e OAIS, uma série de Estudos de Casos e Modelos, assim como uma seção sobre Aplicação da Estrutura às Metodologias de Auditoria.

Essa Estrutura será útil em futuras análises e(ou) auditorias de repositórios que realizam preservação digital de formas altamente distribuídas.

REFERÊNCIAS

CCSDS - CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS. **Producer-Archive Interface Methodology Abstract Standard**, Magenta Book. Washington, DC: CCSDS, May 2004.

CCSDS - CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS. **Reference Model for an Open Archival Information System (OAIS)**. Washington, DC: CCSDS, 2002.

CRL; OCLC. CENTER FOR RESEARCH LIBRARIES; ONLINE COMPUTER LIBRARY CENTER. **Trustworthy Repositories Audit Criteria: Checklist and Certification v1.0**, 2007.

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, 2003.

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

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.

LITTMAN, J. A Technical Approach and Distributed Model for Validation of Digital Objects. **D-Lib Magazine**, v. 12, n. 5, 2006. Disponível em: <http://www.dlib.org/dlib/may06/littman/05littman.html> . 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. **A Holistic Approach to Bit Preservation**. Tese. Doutorado. Copenhagen University, 2011.

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 Bit-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 – Proceedings 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 – Proceedings 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.