



FUNDAÇÃO OSWALDO CRUZ

Plano de Preservação Digital da VideoSaúde - Distribuidora (VSD)

Rio de Janeiro, dezembro de 2020

Fundação Oswaldo Cruz

Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT)

VideoSaúde - Distribuidora da Fiocruz

Plano de Preservação Digital da VideoSaúde - Distribuidora (VSD)

Versão: 1.1

Autores:

Marco Dreer

João Guilherme Nogueira Machado

Eliane Pontes

Colaboradores:

Cleomar Huche Lopes

Karina Veras Praxedes dos Santos Gagliano

Sumário

1.	INTRODUÇÃO	4
2.	CENÁRIO INSTITUCIONAL.....	4
2.1.	ESTRUTURA ORGANIZACIONAL.....	4
2.2.	POLÍTICAS DE PRESERVAÇÃO RELACIONADAS.....	5
2.3.	ASPECTOS LEGAIS.....	5
3.	DESCRIÇÃO DO ACERVO.....	5
3.1.	IDENTIFICAÇÃO E ESCOPO	5
3.2.	ASPECTOS ORGANIZACIONAIS	7
3.3.	ASPECTOS TÉCNICOS.....	8
3.4.	REQUISITOS DOS USUÁRIOS.....	9
3.5.	PRÁTICAS CORRENTES DE PRESERVAÇÃO	9
4.	AVALIAÇÃO DE RISCOS	10
5.	PLANO DE PRESERVAÇÃO DIGITAL.....	14
5.1.	ESTRATÉGIA DE PRESERVAÇÃO AVALIADA: ADOÇÃO DE SOFTWARE EM CONSONÂNCIA COM O MODELO OAIS.....	14
5.2.	FUNÇÕES E RESPONSABILIDADES	15
5.3.	CUSTOS E FONTES DE FINANCIAMENTO	16
5.4.	REVISÃO DO PLANO DE PRESERVAÇÃO DIGITAL.....	16
6.	DETALHAMENTO DO PLANO DE PRESERVAÇÃO DIGITAL	16

1. INTRODUÇÃO

O presente documento integra o Programa de Preservação Digital de Acervos da Fiocruz e tem como objetivo estabelecer diretrizes para a preservação digital do acervo audiovisual da VideoSaúde - Distribuidora (VSD), em conformidade com a atribuição de responsabilidade dada à Distribuidora como mencionado na Política de Preservação dos Acervos Científicos e Culturais da Fiocruz¹.

Este plano tem como objetivo apontar estratégias que possam propiciar a preservação dos documentos audiovisuais digitais da VSD por um longo prazo e ao mesmo tempo assegurar uma presunção de autenticidade aos mesmos, por meio da utilização de operações técnicas e ferramentas específicas.

2. CENÁRIO INSTITUCIONAL

2.1. ESTRUTURA ORGANIZACIONAL

A VideoSaúde - Distribuidora (VSD) é um serviço vinculado ao Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (Icict), unidade técnico-científica da Fundação Oswaldo Cruz (Fiocruz). Entre as suas atribuições, destaca-se a veiculação de conteúdo audiovisual no campo da saúde coletiva em emissoras públicas, educativas e comunitárias. Atualmente exibe sua programação nos seguintes canais: TV UFES, TV Floripa, UFPR TV, TV Câmara, TVE Bahia e Canal Saúde.

Com o objetivo de contribuir para o fortalecimento do Sistema Único de Saúde (SUS) e para a melhoria das condições de vida e saúde da população brasileira, a Distribuidora é um espaço de guarda, produção, captação, preservação, pesquisa, fomento, distribuição e circulação de material audiovisual em saúde produzido pela Fiocruz, por outras instituições e por produtores independentes.

Atualmente, a VSD conta com 21 colaboradores, entre servidores públicos, estagiários e bolsistas, que apoiam atividades ligadas a três grandes áreas: Acervo, Produção e Distribuição.

A VSD recebe parte de seu financiamento através de verba oriunda do orçamento da Fiocruz e repassada para o Icict, além de verbas advindas de captação externa.

¹Disponível em: <https://portal.fiocruz.br/documento/politica-de-preservacao-dos-acervos-cientificos-e-culturais-da-fiocruz>. Acesso em: 14 dez. 2020.

2.2. POLÍTICAS DE PRESERVAÇÃO RELACIONADAS

A VSD vem fortalecendo a cultura institucional e a conscientização da sua equipe e da direção do Icict sobre a importância do desenvolvimento de estratégias de preservação digital do seu acervo, amparadas pelas diferentes políticas institucionais da Fiocruz, que se relacionam diretamente com a questão da preservação do patrimônio científico e cultural da fundação e do acesso aberto ao conhecimento produzido pela instituição, como a Política de Preservação dos Acervos Científicos e Culturais, a Política de Acesso Aberto ao Conhecimento² e a Política de Memória Institucional³.

2.3. ASPECTOS LEGAIS

As obras custodiadas no acervo da VSD seguem os preceitos da Política de Acesso Aberto ao Conhecimento, que torna público o que é produzido na instituição, como forma de garantir o acesso público, gratuito e aberto ao conteúdo de toda obra intelectual produzida pela Fiocruz. Os documentos audiovisuais são obra coletiva que se caracteriza pela fusão de mais de um produto criativo para formar uma obra nova⁴.

A VSD utiliza o Termo de Cessão de Direitos Autorais Geral e o Termo de Cessão de Direitos Autorais Obra Futura, instrumentos elaborados pela Fiocruz e disponíveis na Intranet da instituição. O TC Geral relaciona-se com as produções que são incorporadas na Coleção; e o TC Obra Futura é aplicado às produções da instituição.

3. DESCRIÇÃO DO ACERVO

3.1. IDENTIFICAÇÃO E ESCOPO

O acervo da VSD foi iniciado em 1988, a partir da determinação de seus organizadores em reunir e divulgar vídeos com a temática da saúde pública. O propósito fundamental é fazer circular conteúdo audiovisual entre os diversos públicos interessados, tanto pesquisadores e estudantes quanto a população de um modo geral. Devido aos seus fundamentos históricos, foram sendo constituídos dois conjuntos documentais: um arquivístico, ou seja, o conjunto

² Disponível em: https://portal.fiocruz.br/sites/portal.fiocruz.br/files/documentos/portaria_-_politica_de_acesso_aberto_ao_conhecimento_na_fiocruz.pdf. Acesso em: 14 dez. 2020.

³ Disponível em: <https://portal.fiocruz.br/documento/politica-de-memoria-institucional-da-fiocruz>. Acesso em: 14 dez. 2020.

⁴ Lei 9610/98, art. 5º, VII, h: "coletiva - a criada por iniciativa, organização e responsabilidade de uma pessoa física ou jurídica, que a publica sob seu nome ou marca e que é constituída pela participação de diferentes autores, cujas contribuições se fundem numa criação autônoma".

documental produzido e/ou coproduzido pela Distribuidora, e uma coleção temática de produções externas à Fiocruz.

O sistema de informação do acervo utilizado é o Banco de Recursos Audiovisuais em Saúde (BRAVS)⁵, no qual se encontram registrados e catalogados os títulos que compõem o acervo da VSD. Estes registros fazem referência especialmente aos documentos analógicos, incluindo materiais editados e não editados de diversos formatos, além de séries de diferentes proveniências.

Com o início do tratamento dos objetos digitais, o conjunto arquivístico foi estruturado em quatro séries, sendo elas:

- 1) Programas VideoSaúde: vídeos da coleção, que são selecionados e transformados em programa de televisão;
- 2) Selo Fiocruz Vídeo: produções fomentadas pela Fiocruz por meio de edital público;
- 3) Projetos: materiais produzidos pelo setor relativos às variadas produções;
- 4) Eventos: gravações de eventos técnicos e/ou científicos, editados ou não.

Os formatos de arquivo de vídeo presentes no acervo audiovisual digital são em sua maioria proprietários (incluindo .mov, .m2t, .m2ts, .mp4, .avi e .avchd), nativamente gerados pelas filmadoras de vídeo utilizadas pela VSD, que não oferecem a opção de formatos abertos. Os formatos de arquivo, também conhecidos como contêineres, funcionam como invólucros para os codificadores/decodificadores (*codecs*) dos sinais de vídeo e áudio que compõem os documentos audiovisuais digitais.

Quanto aos formatos de *codec* de vídeo e de áudio mais utilizados atualmente nos fluxos de trabalho da Distribuidora, encontram-se o ProRes e o H.264, para vídeo); e o ACC, para áudio).

O acervo audiovisual digital em questão tem o tamanho aproximado de 80 Terabytes (TB), sendo que a sua previsão de crescimento anual é de aproximadamente 15 TB. Os objetos digitais que compõem o acervo estão armazenados tanto em servidor local quanto em discos rígidos externos e fitas de dados do tipo LTO-6 (Linear Tape-Open, 6ª geração).

Quanto à nomenclatura dos arquivos digitais, é adotado um padrão que segue a estrutura de quadro de arranjo definido pela VSD em parceria com o Sistema de Gestão de Documentos e Arquivos da Fiocruz (Sigda).

⁵ Sistema de catalogação do acervo da VSD desenvolvido pelo Ctic/Icict. Disponível em: <https://bravs.icict.fiocruz.br/>

Quadro 1 - Inventário parcial do acervo audiovisual digital da VSD

Principais formatos analógicos		Estrutura arquivística digital
Formato	Quantitativo	Volume estimado total nato digital: 80 TB
VHS	2.185	<u>Volume tratado*: 2.32 TB:</u> 23 dossiês – série Eventos (não editado); 30 dossiês – série Eventos (editado) 08 dossiês – série Programas VIDEOSAÚDE Cópias de segurança (Backup) – LTO-6 *Considera-se volume tratado o material que já está sendo identificado e submetido aos novos fluxos de procedimento técnico.
SVHS	168	
BETACAM	1.109	
DVCAM	1.461	
DVD	2.351	
MINIDV	1.501	

3.2. ASPECTOS ORGANIZACIONAIS

No que se refere às boas práticas de preservação digital, tendo em vista a gestão do acervo, a Distribuidora vem remodelando seus fluxos, ações e procedimentos de produção, armazenamento e acesso ao conteúdo intelectual. Vale ressaltar o exponencial crescimento do acervo, sobretudo dos vídeos nato digitais.

Em relação aos recursos financeiros e às estratégias ligadas diretamente aos novos fluxos de trabalho voltados para a preservação digital, os impactos imediatos sentidos no cotidiano da Distribuidora se relacionam com:

- 1) Necessidade de infraestrutura tecnológica capaz de atender à demanda por armazenamento dos documentos digitais, sejam eles nato digitais ou oriundos de

- processos de digitalização, incluindo a implantação de um repositório digital de preservação e um laboratório local de digitalização;
- 2) Indispensabilidade de recursos humanos e serviços relacionados à instalação, manutenção e reparo de *hardware* e *software* que integrem a cadeia de ações e procedimentos que estão em fase de testes e implementações;
 - 3) Necessidade de infraestrutura capaz de absorver o material disperso em HDs externos que se acumulam enquanto aguardam espaço para armazenamento e tratamento técnico arquivístico em conformidade com os novos fluxos;
 - 4) Contratos de suporte e/ou manutenção de *hardware* e *softwares*, assim como contratos com empresa de terceirização de recursos humanos, impactam significativamente as atividades desenvolvidas no âmbito das três grandes áreas de atividades da VSD.

3.3. ASPECTOS TÉCNICOS

A VSD conta com um servidor local com estrutura de diretórios baseada em quadro de arranjo em consonância com o modelo hierárquico da Norma Internacional de Descrição Arquivística - ISAD(G).

O servidor está fisicamente instalado nas próprias dependências da VSD, próximo ao campus principal da Fiocruz e localizado na cidade do Rio de Janeiro.

Os objetos digitais audiovisuais da VSD requerem *softwares* que possuam ou comportem os *codecs* de vídeo e de áudio necessários para a leitura dos seus formatos de arquivo específicos. Como os arquivos digitais de vídeo possuem características mais complexas em comparação a outros arquivos digitais, envolvendo a presença de *codecs* específicos, esse monitoramento de *softwares* e *codecs* é um fator de grande relevância para a preservação de longo prazo das informações digitais.

Em seu fluxo de trabalho, a VSD faz uso dos seguintes tipos de metadados: administrativos, descritivos e de preservação. Para os metadados descritivos, estão sendo utilizados elementos de padrões de metadados já consolidados, como o Dublin Core, amplamente aceito e utilizado, e o PBCore, específico para documentos audiovisuais. Já os metadados administrativos (técnicos) são extraídos em elementos PBCore por meio do *software* de código aberto MediaInfo. Atualmente está sendo avaliada e testada a utilização de um padrão de metadados que atenda às especificidades do vídeo, para a complementação do padrão PREMIS no registro do histórico dos processos de reformatação. Esse perfil de metadados de digitalização de documentos audiovisuais visa propiciar a manutenção da identidade e integridade dos representantes digitais gerados, atribuindo, assim, uma maior presunção de autenticidade aos mesmos.

Em meados de 2018 se iniciaram testes com o *software* Archivematica, sistema gratuito de preservação digital de código aberto, por meio do qual determinados arquivos digitais de

vídeo vêm sendo admitidos a fim de se observar a eficiência da plataforma para o contexto do acervo da VSD, em particular no que se refere à sua preservação de longo prazo.

3.4. REQUISITOS DOS USUÁRIOS

O acervo da VSD está disponível para usuários cadastrados, entre eles organismos e instituições do Sistema Único de Saúde (SUS), entidades privadas, escolas, estudantes, organizações não governamentais e comunitárias, além de usuários individuais.

Os usuários internos do acervo digital da VSD incluem os colaboradores, que são responsáveis pela produção e gestão dos objetos digitais, além daqueles dedicados à sua preservação. Os usuários necessitam de uma solução de preservação digital que:

- Seja segura, a fim de garantir a integridade do acervo digital: os dispositivos de armazenamento deverão ter seus dados acessados somente a usuários autorizados.
- Seja confiável: deverá haver um fluxo compreensível e confiável dos dados digitais, desde o descarregamento do conteúdo dos cartões de memória até o armazenamento nas diferentes mídias utilizadas.
- Otimize e agilize os processos de edição de conteúdo, sobretudo para os usuários ligados à produção/edição. Qualquer solução proposta deve garantir que todo o conteúdo demandado, por meio de um planejamento prévio e mínimo, esteja disponível para as máquinas no local (servidor local VSD).
- Automatize a criação de pacotes de informação de submissão (SIP).
- Possibilite interoperar o BRAVS com o Archivematica (disseminação DIP).
- Formalize um fluxo de trabalho documentado que contemple as ações e processos voltados para a preservação digital em conformidade com normas e recomendações de melhores práticas.

3.5. PRÁTICAS CORRENTES DE PRESERVAÇÃO

A VSD realiza *backup* dos seus arquivos digitais de maneira regular, com periodicidade semanal, sendo geradas duas cópias de segurança, além do arquivo original, para cada objeto digital, em dois suportes distintos (disco rígido e fita magnética LTO). O servidor local trabalha com o armazenamento em RAID (sigla em inglês para *Redundant Array of Inexpensive Disks*) e os dois *backups* são realizados em fitas LTO 6, utilizando estratégia de dispersão geográfica. Quanto ao controle ambiental, há verificação e registro diário de temperatura e umidade relativa tanto na área do acervo analógico quanto na área destinada ao servidor de arquivos digitais da VSD.

Outra ação de preservação digital adotada pela VSD envolve a monitoração dos processos de obsolescência tanto dos formatos de arquivos quanto dos suportes de armazenamento. Portanto, quando necessárias, são realizadas migrações de suportes sob risco de obsolescência.

Baseando-se nas melhores práticas de preservação digital orientadas pelo modelo OAIS, a Distribuidora trabalha com o conceito de pacotes de informação. Nesse sentido, dois procedimentos se fazem fundamentais no fluxo de processamento técnico dos materiais: o empacotamento dos arquivos de acordo com padrão BagIt⁶ e a aferição de integridade e fixidez dos mesmos através da verificação de checksum. Para essas ações são utilizadas, respectivamente, as ferramentas gratuitas e de código aberto, Exactly e Fixity.

Há um controle de acesso aos objetos digitais por meio de autorizações concedidas por diferentes níveis de acesso (leitura, edição, exclusão etc).

4. AVALIAÇÃO DE RISCOS

A fim de fornecer as bases para a elaboração de uma estratégia de preservação digital, é recomendável avaliar previamente os riscos que possam impactar negativamente a gestão e a preservação do acervo audiovisual da VSD, além de indicar ações de tratamento possíveis. Os Quadros 1 e 2, apresentados a seguir, baseiam-se em modelo e critérios apresentados no *Guia de gestão de riscos da Fiocruz*⁷; por sua vez, as áreas funcionais indicadas no Quadro 2 baseiam-se no *Levels of Digital Preservation*⁸ (Níveis de Preservação Digital) – conjunto hierárquico de recomendações desenvolvido pela Library of Congress.

A avaliação preliminar apresentada no Quadro 2 indica níveis de risco de médio a alto, associados à preservação dos documentos digitais da coleção em questão. Os níveis mais acentuados de risco, que obrigatoriamente requerem ação mais imediata, envolvem três fatores: 1) perda ou corrupção dos arquivos digitais; 2) alteração inesperada ou deliberada dos arquivos digitais; e 3) acesso não autorizado aos arquivos digitais. Já os riscos de nível médio giram em torno de dois fatores: 1) metadados insuficientes; e 2) obsolescência de formatos de arquivo e dispositivos de armazenamento. No mesmo quadro são apontados também os possíveis impactos associados a cada um dos riscos, os controles existentes para reduzir o seu impacto, além das possíveis ações de tratamento para a sua mitigação.

⁶ BagIt é um conjunto de convenções de layout de arquivo hierárquico projetado para suportar armazenamento e transferência de conteúdo digital arbitrário.

⁷ Disponível em: https://www.arca.fiocruz.br/bitstream/icict/39359/2/guia_de_gestao_de_riscos_da_fiocruz.pdf; acesso em 16 dez. 2020.

⁸ Disponível em: <https://ndsa.org/publications/levels-of-digital-preservation/>. acesso em: 16 dez. 2020.

Quadro 1 – Níveis de Risco

IMPACTO	Muito Alto 10	10 RM	20 RM	50 RA	80 RE	100 RE
	Alto 8	8 RB	16 RM	40 RA	64 RA	80 RE
	Médio 5	5 RB	10 RM	25 RM	40 RA	50 RA
	Baixo 2	2 RB	4 RB	10 RM	16 RM	20 RM
	Muito Baixo 1	1 RB	2 RB	5 RB	8 RB	10 RM
		Muito Baixa 1	Baixa 2	Média 5	Alta 8	Muito Alta 10
PROBABILIDADE						

- RB Risco Baixo
- RM Risco Médio
- RA Risco Alto
- RE Risco Muito Alto/ Estratégico

Quadro 2 – Planilha de identificação e avaliação de riscos

Área Funcional	Risco	Consequências/Impactos	Nível de Risco	Controles Existentes	Ações de Tratamento
Armazenamento	Perda ou corrupção dos arquivos digitais	Incapacidade de ler, acessar e recuperar conteúdos.	ALTO Probabilidade (5) X Impacto (8) = 40	Os arquivos digitais da coleção são mantidos em diferentes tipos de armazenamento (HDs externos, servidor e fitas LTO), havendo redundância e dispersão geográfica (cópia de segurança localizada na sala-cofre do campus de Manginhos).	Documentar as mídias de armazenamento utilizadas, a fim de assegurar sua longevidade, indicando os recursos e requisitos necessários para o seu funcionamento.
Integridade	Alteração inesperada ou deliberada dos arquivos digitais	Perda de fixidez dos arquivos digitais, comprometendo a sua integridade. A capacidade de manutenção da integridade é um dos elementos essenciais da preservação digital.	ALTO Probabilidade (8) X Impacto (8) = 64	Só há verificações de fixidez e de presença de vírus nos arquivos digitais que estão armazenados no servidor da STI. Não há procedimentos desse tipo nos dispositivos localizados na VSD.	Geração de soma de verificação (<i>checksum</i>) de todos os arquivos digitais, além de sua verificação a cada seis meses, no mínimo. Os valores de <i>checksum</i> deverão ser gerados e verificados em todos os dispositivos onde estão armazenados os arquivos digitais. Além disso, realizar verificação de vírus, rotineiramente, em todos os dispositivos.
Controle	Acesso não autorizado aos arquivos digitais	Corrupção, alteração ou exclusão dos arquivos digitais.	ALTO Probabilidade (5) X Impacto (8) = 40	O acesso aos arquivos digitais armazenados nos HDs externos e no servidor local é restrito à equipe técnica autorizada da VSD, mas não é formalmente documentado.	Documentar todos os profissionais que terão autorização para ler, editar e mover os arquivos digitais da coleção.

Metadados	Metadados insuficientes	Impossibilidade de descrever, compreender, encontrar e recuperar o conteúdo adequadamente. Metadados consistentes são cruciais para prover presunção de autenticidade aos conteúdos, incluindo informações sobre as ações de preservação realizadas ao longo do ciclo de vida dos arquivos digitais.	MÉDIO Probabilidade (5) X Impacto (5) = 25	Alguns metadados descritivos e técnicos estão de acordo com os metadados especificados no Manual de Digitalização da Fiocruz ⁹ , mas são insuficientes para descrever plenamente os conteúdos.	Geração de metadados técnicos para todos os documentos digitais da coleção.
Conteúdo	Obsolescência de formatos de arquivo e dispositivos de armazenamento	Incapacidade de reprodução e leitura dos arquivos digitais, comprometendo o acesso aos seus conteúdos.	MÉDIO Probabilidade (5) X Impacto (5) = 25	Uso do formato MKV/FFV1 (contêiner/codec) como representante digital oriundo de materiais digitalizados e monitoramento dos processos de obsolescência dos formatos de arquivo, sobretudo nato digitais.	Elaborar estratégia de normalização de formatos de arquivo, voltados para a preservação.

⁹ Disponível em: <https://portal.fiocruz.br/noticia/manual-de-digitalizacao-da-fiocruz-2019>. Acesso em: 22 dez. 2020.

5. PLANO DE PRESERVAÇÃO DIGITAL

5.1. **ESTRATÉGIA DE PRESERVAÇÃO AVALIADA: ADOÇÃO DE SOFTWARE EM CONSONÂNCIA COM O MODELO OAIS**

A VSD vem remodelando seus fluxos e ações de produção, processamento técnico, arquivamento e acesso ao seu acervo audiovisual com vistas a garantir o cumprimento das mais atualizadas e consolidadas técnicas de preservação digital em consonância com recomendações do modelo de referência OAIS/ISO 14721:2012. Esta estratégia de preservação digital se concentra no uso de *software* voltado para fornecer uma solução de preservação digital que esteja compatível com os padrões do Open Archival Information System (OAIS). O *software* Archivemática, desenvolvido pela empresa canadense Artefactual, vem recebendo adoção significativa por diversas instituições de guarda, tanto no Brasil como no exterior, por ser solução tecnológica que se aproxima dos requisitos de preservação digital estabelecidos pelo modelo OAIS. No entanto, a adoção do *software* requer instalação, configuração, customização e operação complexas, acarretando em considerável tempo de trabalho e custos.

Por meio de um conjunto estruturado de fluxos de trabalho, a VSD vem adotando a estratégia de criação de pacotes de informação de submissão (SIPs) para futura admissão no *software* de preservação (Archivemática). Os fluxos de trabalho se relacionam desde o momento de produção dos documentos até o seu arquivamento permanente e disponibilização para acesso público e aberto aos vídeos da VSD. Esses fluxos se integram a partir da captura e registro de imagens em movimento que irão compor os documentos, assim como na coleta e geração de metadados descritivos sobre o seu conteúdo, até chegarem ao estágio de ingestão via descarregamento ou transferência no servidor local, seguido do processamento técnico para arquivamento. Trata-se de um arquivamento local de SIPs para posterior ingestão no Archivemática, de modo que os SIPs criados somente se tornarão AIPs quando forem admitidos no *software*.

Todas as ferramentas (*softwares*) utilizadas atualmente pela VSD são gratuitas e de código aberto, servindo a diferentes propósitos:

DROID: descrição pormenorizada dos formatos de arquivo digitais, a fim de auxiliar no inventário;

Exactly: ferramenta gratuita e de código aberta utilizada para empacotamento no padrão BagIt, desenvolvido pela Library of Congress (LOC) e reconhecido pela Internet Engineering Task Force (IETF).

Fixity: Análise de integridade e fixidez dos objetos digitais, por meio da verificação de valor de *checksum* (do tipo MD5) com registro de histórico e relatório das verificações;

Libreoffice CALC: construção das planilhas de metadados descritivos, em formato .csv.

MediaInfo: extração de metadados técnicos embutidos;

Robocopy: verificação de integridade e manutenção de metadados técnicos;

A VSD iniciou testes com o Archivematica em meados de 2018, de modo que arquivos digitais de vídeo selecionados estão sendo submetidos à plataforma a fim de se observar a plena adequação do *software* para o contexto do acervo da VSD e sobretudo como solução eficiente de preservação digital de longo prazo.

A estratégia de preservação também deverá contemplar os requisitos de redundância e dispersão geográfica, essenciais para qualquer ação de preservação digital. Nesse sentido, nota-se a necessidade do compartilhamento de competência técnica e estrutural com a COGETIC no que tange ao armazenamento e às estratégias de gestão de riscos. Uma possível estratégia é a adoção de redundância do servidor local da VSD pela COGETIC e a dispersão da replicação *offline* em fitas LTO nas áreas de guarda da Biblioteca de Manguinhos, localizada no campus principal da Fiocruz.

Vantagens da estratégia avaliada:

- Uso de *software* (Archivematica) que se aproxima mais dos requisitos da norma OAIS.
- Maior automação das ações de preservação digital, que são contempladas por uma única ferramenta.
- Menos tempo dispensado nas ações de preservação digital, uma vez que estão centralizadas numa única ferramenta, mais automatizada.

Desvantagens da estratégia avaliada:

1. Processos de instalação, configuração, customização do *software* apresentam grande complexidade, acarretando custos diretos e indiretos.
2. Solução que requer treinamento intensivo por parte dos usuários da ferramenta, dada a sua complexidade.
3. Apesar de ser um *software* aberto e de adoção cada vez maior em instituições de guarda digital, não há como garantir a plena sustentabilidade da ferramenta caso a empresa desenvolvedora venha a encerrar suas atividades.

5.2. FUNÇÕES E RESPONSABILIDADES

Para a implementação do plano, pelo menos seis profissionais estarão mobilizados: dois diretamente envolvidos na preservação, um no suporte à preservação e TI, e três na área de produção/edição. Além desses, haverá a necessidade de contratação de profissionais com formação na área de arquivo e preservação digital.

Entre as atividades pelas quais ficarão responsáveis os profissionais da VSD e da TI do Icict no contexto do plano, incluem: verificação de vírus, análise e transferência de mídias, inserção e gestão de metadados, conferência e controle de qualidade, geração de *backup*, entre outros procedimentos.

A COGETIC deverá ficar responsável pela replicação periódica dos dados digitais armazenados no servidor local da VSD para servidor próprio, incluindo a gestão e preservação dos mesmos.

5.3. CUSTOS E FONTES DE FINANCIAMENTO

Os custos necessários para a implementação do plano de preservação digital deverão ser levantados conjuntamente entre os profissionais da VSD, a TI do Icict e a COGETIC. Os custos deverão contemplar tanto infraestrutura tecnológica quanto recursos humanos, além da previsão de investimentos necessários na atualização contínua dos recursos preservados (incluindo *hardware*, *software* e sistemas), de modo a dirimir os processos de obsolescência tecnológica.

As principais fontes de financiamento para o plano devem envolver sobretudo os editais de fomento dedicados ao campo do patrimônio, além de recursos advindos do orçamento da Fiocruz.

5.4. REVISÃO DO PLANO DE PRESERVAÇÃO DIGITAL

Entre os eventos que poderão desencadear a revisão do Plano de Preservação Digital da VSD, pode-se destacar:

- Alterações no orçamento previsto para a preservação digital, podendo gerar alterações de investimentos e custos relacionados;
- Alterações na tecnologia adotada pela VSD em uma ou mais etapas da cadeia da preservação digital, como, por exemplo, a interrupção do uso de fitas do tipo LTO como estratégia de *backup* ou a adoção de outras ferramentas de preservação digital;
- Alterações no número de profissionais envolvidos diretamente com a preservação digital da VSD, o que poderá acarretar mudanças nos fluxos de trabalho.

6. DETALHAMENTO DO PLANO DE PRESERVAÇÃO DIGITAL

O Plano de Preservação Digital envolve um conjunto de fluxos de trabalho, descritos a seguir e ilustrados no diagrama abaixo.

Após o descarregamento ou transferência dos conteúdos advindos de cartões de câmeras digitais, *pen drive*, FTP etc., os vídeos passam pelo processo de verificação de vírus e tramitação por meio da ferramenta Robocopy (para manutenção dos atributos de metadados gerados no ato de gravação do “original”) para garantia de presunção de autenticidade dos conteúdos. Todos os conteúdos admitidos no servidor local passam por checagem de vírus – sejam eles arquivísticos, sejam pertencentes à coleção temática.

A partir da ingestão dos conteúdos, o responsável pelo tratamento técnico irá realizar, após a criação dos pacotes de informação, atividades de nomeação de arquivos, que em seguida serão armazenados em dossiês criados de maneira sequencial e em conformidade com a estrutura do quadro de arranjo definido. A estrutura de dossiê será orientada pelo modelo padrão de pasta “dossiê = título”, contendo duas subpastas: “objects”, onde são depositados os arquivos a serem preservados; e “metadata”, que abriga uma planilha com os metadados dos objetos depositados na primeira pasta. Caso haja necessidade, outras subpastas dentro do dossiê correspondente poderão ser criadas para atenderem a sua finalidade.

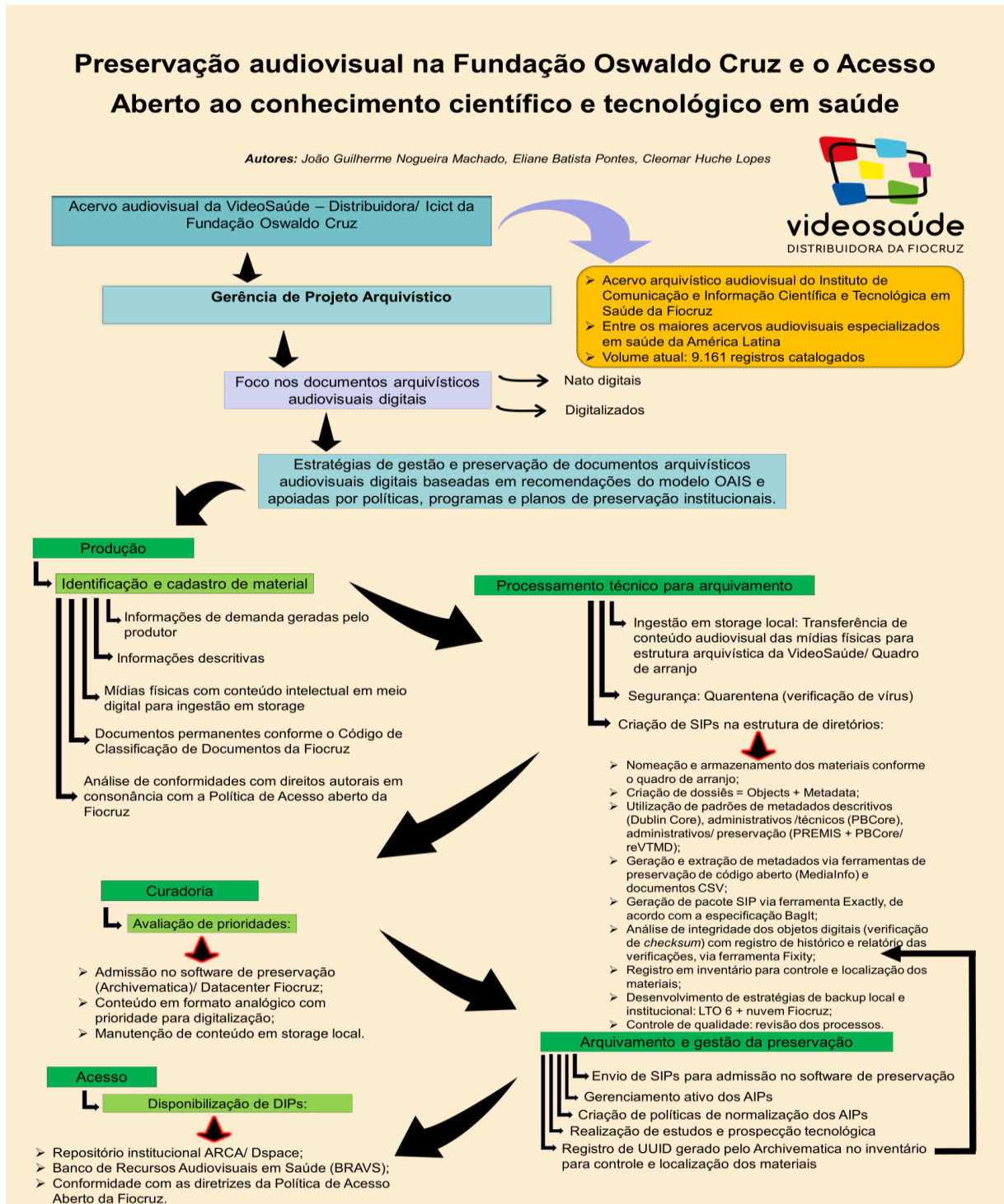
A geração e/ou extração de metadados descritivos, técnicos e de preservação, assim como a criação e verificação dos valores de *checksum* se dará através de ferramentas de preservação digital gratuitas e de código aberto. Os metadados descritivos são gerados em planilhas, por meio do *software* Libreoffice CALC; enquanto que os metadados técnicos embutidos são extraídos através do MediaInfo. Finalmente, os dossiês serão empacotados em BagIt via ferramenta Exactly e serão monitorados quanto sua integridade e fixidez ao longo do tempo através da ferramenta Fixity até o momento de sua admissão no Archivematica.

Após os processos de tratamento técnico, é realizada uma revisão final do pacote com objetivo de manter um controle de qualidade que atenda aos requisitos e recomendações do campo da preservação digital.

O próximo passo será validar o SIP de acordo com as diretrizes do OAIS e trabalhar com a metainformação gerada e acrescentada na gestão ativa dos SIPs, para posterior geração de AIPs por meio do repositório digital. As informações coletadas e geradas na elaboração dos SIPs e na sua validação enquanto AIPs também serão registradas em planilha de inventário para localização e recuperação dos pacotes – seja na estrutura de *storage* local, seja em um repositório de preservação digital.

Para a difusão, serão normalizados arquivos em formato de visualização para acesso no repositório institucional ARCA da Fiocruz (baseado na plataforma DSpace), assim como através do Banco de Recursos Audiovisuais em Saúde (BRAVS) da própria VSD/Icict. Para a criação de pacotes de informação de disseminação (DIP - Dissemination Information Package) no âmbito do BRAVS, estudos preliminares e que viabilizem a interoperabilidade com o sistema estão sendo realizados em parceria com o CTIC/Icict.

Figura 1 - Diagrama dos fluxos de trabalho dos objetos digitais audiovisuais da VSD



Av. Brasil, 4.365 - Pavilhão Haity Moussatché - Manguinhos, Rio de Janeiro | CEP: 21040-900 | Tel.: (+55 21) 3865-3131 | Fax.: (+55 21) 2270-2668 | www.icict.fiocruz.br

PRODUÇÃO:

- Solicitação/demanda via abertura de registro/produção;
- Informações descritivas sobre o documento:
 - informações fornecidas pelo demandante/produtor através da plataforma/formulário de solicitação;
 - informações de confirmação da demanda + informações descritivas codificadas para Dublin Core em formatos JSON e CSV (codificação em UTF-8) geradas no ato de preenchimento e confirmação (“ok”) do formulário de solicitação *online*,
- No ato da confirmação (“ok”), após o preenchimento do formulário pelo demandante/produtor, um email deverá ser enviado para um receptor/destinatário contendo os arquivos com os metadados Dublin Core em JSON e CSV. Esses arquivos serão incluídos na subpasta “metadata” em um dossiê (título do documento) com a nomeação correspondente ao número de sua respectiva série, de acordo com o quadro de arranjo estabelecido:

ex: BR_RJVSD_09_01_01_003_01_0001de0001

- O dossiê será criado em um ambiente “intermediário”, aguardando a chegada dos arquivos digitais que comporão a subpasta “objects” do respectivo dossiê, para o posterior processamento técnico e geração do SIP. Na criação do SIP serão utilizadas as ferramentas para geração e extração de metadados complementares, empacotamento e registro de histórico e relatório de verificação de *checksum* (fixidez e integridade): MedialInfo, Exactly, Fixity, Robocopy e Libreoffice CALC.

PROCESSAMENTO TÉCNICO:

O processamento técnico se dará a partir do momento do recebimento e descarregamento do conteúdo dos cartões e/ou mídias de armazenamento, contendo arquivos que comporão a subpasta “objects” do dossiê correspondente no ambiente “intermediário” (dossiê já aberto e com os metadados descritivos gerados em, JSON e CSV (codificado em UTF-8)).

No ato do descarregamento dos arquivos em seus respectivos dossiês, será realizada uma verificação de antivírus. Após esse processo, a ferramenta Robocopy será utilizada para a verificação de integridade e manutenção de metadados técnicos originais. Esses procedimentos deverão ser registrados na planilha de controle/inventário dos materiais. Após esse primeiro passo, os cartões são liberados para reutilização.

Com os arquivos descarregados, o próximo passo do processamento técnico será a extração de metadados técnicos, via MedialInfo, do arquivo ou lote de arquivos que componham a subpasta “objects”. Os metadados gerados no MedialInfo serão exportados no padrão PBCore em formato XML, nomeando os arquivos (em letras minúsculas) seguindo o padrão “metadata_xxx.xml” e inserindo-os na subpasta “metadata” do respectivo dossiê, juntamente com os metadados em padrão Dublin Core - DC (“xxx” pode ser pbcore ou dc).

Para a conclusão da última etapa do processamento técnico e da criação dos SIPs (através da ferramenta de empacotamento Exactly), será necessária uma verificação dos metadados gerados/extraídos, complementando-os caso seja necessário, além da verificação de nomeação do dossiê e dos arquivos que compõem as suas subpastas.

Após essa verificação, será dado o “ok” para a confirmação do empacotamento do SIP a ser admitido no *software* de preservação, através do preenchimento da planilha de controle/inventário com as informações pertinentes aos processos.

ARQUIVAMENTO:

As ações desta etapa fazem a interseção entre as duas estratégias descritas acima, desde a criação e gerenciamento dos SIPs até a sua posterior admissão no repositório para a geração de AIPs – baseado nas diretrizes do OAIS. As ações seriam as seguintes:

- Admissão dos SIPs no Archivematica com a geração dos AIPs e registro de identificador único universal (UUID) AIP na planilha controle/inventário.
- Definição do planejamento de preservação no Archivematica para conteúdo audiovisual nos microsserviços de:
 - Identificação:
 - Ferramentas
 - Regras
 - Comandos
 - Formatar registro de política:
 - Ferramentas = para a extração de metadados técnicos via MediaInfo (do FITS) no Archivematica, configurar a exportação e exibição dos metadados na “techMD” do METS AIP no padrão PBCore.
 - Caracterização:
 - Regras
 - Comandos
 - Detalhe do Evento:
 - Regras
 - Comandos

- Extração:
 - Regras
 - Comandos
- Normalização:
 - Regras
 - Comandos
- Transcrição:
 - Regras
 - Comandos
- Validação:
 - Regras
 - Comandos
- Verificação:
 - Regras
 - Comandos
- Gerenciamento ativo dos AIP armazenados no Archivematica com suporte dos metadados METS AIP e sua possível leitura via *software* METSFlask (<http://bitarchivist.pythonanywhere.com/>).
- Liberação de espaço em *storage* local após admissão na instância do Archivematica na sala cofre/COGETIC.