



ARQUITETURA CONSERVATIO
ARQUITETURA CONSERVATIO

EMPACOTAMENTO, SUBMISSÃO E DISSEMINAÇÃO DE OBJETOS DIGITAIS E DADOS AO CONSERVATIO

O futuro é digital
e a preservação
é necessária.





RICARDO DE REZENDE FERRAÇO

Governador do Estado do Espírito Santo

MARCELO CALMON DIAS

Secretário de Estado de Gestão e Recursos Humanos

EDMAR MOREIRA CAMATA

Secretário de Estado de Controle e Transparência

MARCELO AZEREDO CORNÉLIO

Instituto de Tecnologia da Informação e Comunicação do Espírito Santo

FABRÍCIO NORONHA

Secretário de Estado da Cultura

CILMAR CESCINETTO FRANCISCETTO Arquivo Público do Estado do
Espírito Santo

Organização

Alex Pereira de Holanda

Juliana Oliveira de Almeida

Jussara Teixeira

Tânia Barbosa Salles Gava

Revisão

Ana Maria dos Santos Latgé

Gabriela De Grande Guerreiro

Diana Vilas Boas Souto Aleixo



FICHA CATALOGRÁFICA

Dados Internacionais de Catalogação na Publicação (CIP)
Ana Carolina Medici Rocha – CRB 1151

E55 Empacotamento, submissão e disseminação de objetos digitais e dados ao CONSERVATIO / organização Alex Holanda, Juliana Oliveira de Almeida, Jussara Teixeira e Tânia Barbosa Salles Gava. – Vitória, ES: APEES ; PRODEST ; UFES ; FAPES, 2026.

1 recurso online (20 p.)

ISBN: 978-65-02-26338-9

1. Preservação digital. 2. Repositórios arquivísticos digitais confiáveis. 3. Modelo OAIS. 4. Objetos digitais. 5. Pacotes de informação. I. Holanda, Alex, org. II. Almeida, Juliana Oliveira de, org. III. Teixeira, Jussara, org. IV. Gava, Tânia Barbosa Salles, org. V. Título.

CDD – 025.84

Copyright © 2026 by Arquivo Público do Estado do Espírito Santo
Rua Sete de Setembro, 414 – Centro, CEP: 29.015-905 - Vitória / Espírito Santo
Tel.: 36366100

E-mail: faleconosco@ape.es.gov.br

Disponível em: <https://proged.es.gov.br/publicacoes>

Permitida a reprodução parcial ou total desde que indicada a fonte



Sumário

1 INTRODUÇÃO	4
1.1 IDENTIFICAÇÃO	4
1.2 OBJETIVO	4
2 CENÁRIO 1– EMPACOTAMENTO POR DEMANDA	6
2.1 DESCRIÇÃO	6
2.2 SEQUÊNCIA DE PASSOS DA SUBMISSÃO	7
2.2.1 Etapas da Primeira Fase de Implementação da Arquitetura Computacional – Submissão de Pacotes de Informação (SIP)	7
2.2.2 Etapas da Segunda Fase de Implementação da Arquitetura Computacional – Gestão da Preservação Digital	11
2.2.3 Auditoria da Submissão	14
2.2.4 Sequência de passos da Auditoria da Submissão	17



1 INTRODUÇÃO

1.1 IDENTIFICAÇÃO

Produto	Arquitetura de preservação digital do projeto PDS
Módulo	ConservatIO
Órgão/Área	Arquivo Público do Estado do Espírito Santo (APEES) / Gerência de Gestão de Documentos (GEDOC)

1.2 OBJETIVO

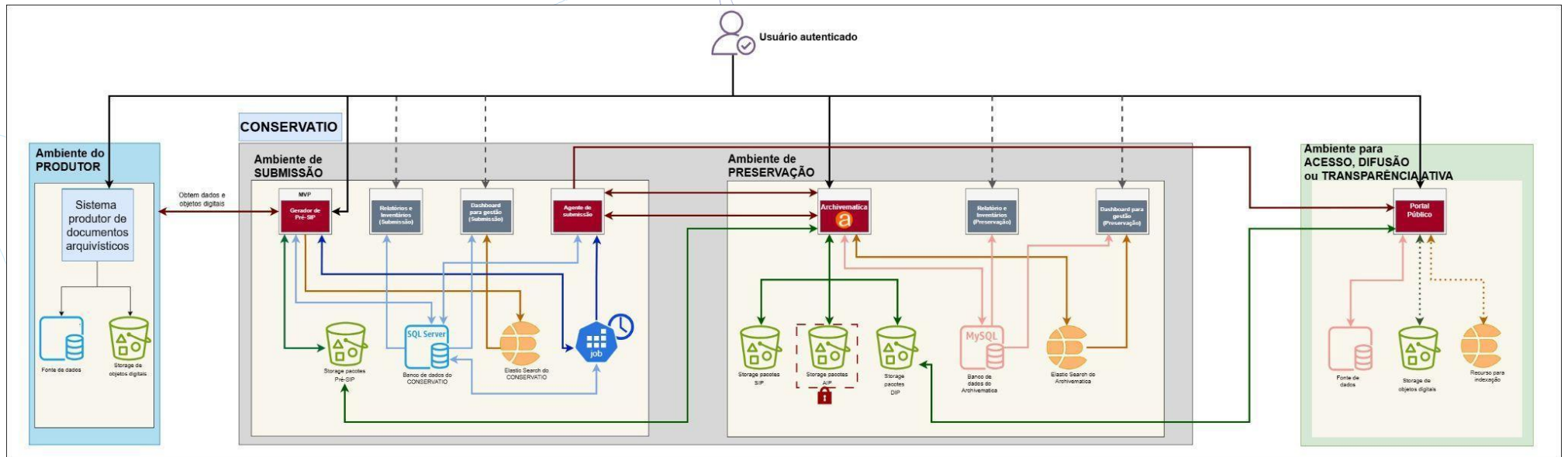
O **ConservatIO** foi concebido como uma solução voltada à preservação digital e ao gerenciamento integrado de documentos arquivísticos, atendendo às demandas de interoperabilidade, de rastreabilidade e de conformidade normativa. A arquitetura proposta busca oferecer uma visão estruturada dos componentes que fazem parte do sistema, destacando sua capacidade de integrar diferentes módulos e tecnologias em um fluxo coeso e auditável.

A descrição da arquitetura do **ConservatIO** tem como objetivo apresentar a forma como os elementos de *software*, os serviços e as camadas de dados se organizam para sustentar as funções essenciais da aplicação. Além disso, busca-se evidenciar as escolhas tecnológicas que viabilizam a escalabilidade, a segurança da informação e a adequação aos princípios arquivísticos de autenticidade, organicidade e confiabilidade.

Com essa perspectiva, a arquitetura fornece não apenas um guia de implementação e de evolução do sistema, mas também um ponto de referência para futuras iniciativas relacionadas à preservação digital, permitindo que a arquitetura **ConservatIO** seja entendido como uma base sólida para soluções institucionais de longo prazo.



Figura 1 - Arquitetura do ConservatIO



Fonte: Elaborado pelos autores.



2 CENÁRIO 1– EMPACOTAMENTO POR DEMANDA

2.1 DESCRIÇÃO

No primeiro cenário, o objetivo é permitir que um usuário autenticado realize todo o ciclo de preservação digital de um objeto, desde o empacotamento inicial até sua disponibilização para acesso. O processo terá início com a seleção, por demanda, de um objeto digital e seus dados, que serão empacotados conforme um dos padrões suportados — Standard, BagIt ou e-ARK —, garantindo a organização, a autenticidade, a confiabilidade e a rastreabilidade dos dados.

Nesse cenário, será implementado ao Archivematica um mecanismo de submissão automatizada de pacotes de informação (pré-SIP), utilizando suas *Application Programming Interface* (API). A submissão será feita com base em um *pipeline* previamente configurado no ambiente do Archivematica, o que garantirá que todos os microsserviços necessários para a ingestão e geração dos pacotes *Submission Information Package* (SIP), *Archival Information Package* (AIP) e *Dissemination Information Package* (DIP) sejam executados.

Com essa integração, busca-se eliminar etapas manuais, garantir consistência na submissão dos pacotes e permitir a validação automática dos pacotes AIP e DIP gerados, conforme as exigências do projeto de preservação digital.

Após o empacotamento, o sistema realizará a transmissão automatizada do pacote para o Archivematica, que executará os microsserviços de ingestão previstos no modelo *Open Archival Information System* (OAIS), tais quais: verificação de integridade, validação de formatos, identificação, normalização e geração dos pacotes de preservação (AIP) e de disseminação (DIP).

Concluída a ingestão, o pacote de disseminação (DIP) será, então, integrado à uma **plataforma de acesso, difusão ou transparência ativa (sistema de disseminação)**, onde os metadados descritivos e o objeto digital serão publicados e disponibilizados para consulta por meio da interface pública.

Esse cenário permite testar e validar toda a cadeia funcional de preservação digital, assegurando a interoperabilidade entre os componentes da arquitetura e o cumprimento dos requisitos definidos no acordo de submissão.



Todo o processo de submissão será registrado em trilha de auditoria pelo "**Gerador de Pré-SIP**" e pelo "**Agente de submissão**", de forma estruturada, garantindo total rastreabilidade e transparência das operações realizadas. Para isso, será mantido um registro detalhado do empacotamento e de cada evento do fluxo do *pipeline*, desde a solicitação inicial de empacotamento até a disseminação do objeto digital.

2.2 SEQUÊNCIA DE PASSOS DA SUBMISSÃO

2.2.1 Etapas da Primeira Fase de Implementação da Arquitetura Computacional – Submissão de Pacotes de Informação (SIP)

1. O **Usuário autenticado** acessa o **ConservatIO ("Gerador de Pré-SIP")** e informa a origem, nome do pacote e o ID do objeto digital a ser preservado. Em seguida, seleciona o tipo de empacotamento a ser aplicado ao objeto digital, dentre as opções disponíveis *Standard*, *BagIt* ou *e-ARK*. Ressalta-se que nesta arquitetura a opção escolhida foi o tipo *Standard*.
2. O "**Gerador de Pré-SIP**" inicia o processo de extração do objeto digital e dos dados associados diretamente do **Ambiente do produtor**. A partir dessas informações, é gerado um pacote no formato Pré-SIP, estruturado conforme o tipo de empacotamento selecionado pelo usuário. O Pré-SIP é armazenado no diretório de transferência previamente configurado no *Storage Service* do Archivematica (**Storage Pacotes Pré-SIP**), tornando-se disponível para o processo de ingestão. Esse registro retorna o local de armazenamento do Pré-SIP, a data e a hora da operação.
3. Durante o processo de empacotamento, as informações essenciais relacionadas à operação são registradas em um **Banco de Dados SQL Server**. Esse registro inclui dados como o identificador do objeto digital, o tipo de empacotamento selecionado (*Standard*), o local de armazenamento do Pré-SIP, além da data e da hora da operação, que são armazenados com finalidade de rastreabilidade, auditoria, controle de versões e suporte à automação do fluxo de preservação digital. O sistema retorna para o "**Gerador de Pré-SIP**" o status do processo.
4. Após o empacotamento e o armazenamento das informações no banco de dados, o conjunto de metadados do objeto digital e do pacote Pré-SIP são enviados para indexação no **Elasticsearch**. Essa indexação inclui dados como o número do processo, o tipo de empacotamento, a identificação dos objetos digitais, o status da submissão, a



data de criação do pacote e as referências ao local de armazenamento. A indexação visa possibilitar a recuperação eficiente e em tempo real dessas informações por meio de buscas textuais, filtros e painéis de visualização, contribuindo para o monitoramento operacional, a auditoria e a integração com outros sistemas da arquitetura de preservação digital.

5. Concluído o processo de empacotamento e de registro das informações, o **módulo "Gerador de Pré-SIP"** aciona, automaticamente, um *job* agendado no *Hangfire*. Esse *job* será responsável por ativar o **ConservatIO (Agente de Submissão)**, que inicia o processo de transferência do pacote Pré-SIP para o Archivematica por meio de chamadas à sua API, retornando o ID do *Job*.
6. O *job* ativa o **Agente de Submissão**, que irá garantir a continuidade do fluxo de preservação digital de forma assíncrona e controlada, permitindo que a submissão seja executada em segundo plano sem impactar a interface do usuário ou as demais operações do sistema.
7. O **Agente de Submissão**, ao ser ativado pelo *job* do *Hangfire*, realiza a comunicação com a **API de Transfer** do Archivematica para iniciar o processo de ingestão do pacote Pré-SIP armazenado no diretório de transferência. Esse agente executa chamadas autenticadas ao *endpoint* de submissão da API (*/api/transfer/start_transfer/*), informando os dados necessários, como o nome do pacote, o tipo de transferência, o caminho relativo no *storage* e a opção automatizada de submissão. Uma vez enviada a solicitação, o agente monitora o retorno da API para validar o aceite do pacote pelo *pipeline* configurado no Archivematica.
 - 7.1. O **Agente de Submissão** inicia o processo de transferência ativando a **API de Transfer** do Archivematica.
 - 7.2. A **API de Transfer** acessa a área de transferência (*Transfer Source Location*) para obter o pacote Pré-SIP informado. Ao final do processo, ela informa o status da transferência.
 - 7.3. O **Agente de Submissão** inicia o processo de ingestão, ativando a **API de Ingest do Archivematica**. A **API de Ingest** executa o processo de ingestão e informa ao agente se a ingestão foi realizada com sucesso.
 - 7.4. Ao longo de todo o fluxo de submissão, o Agente de Submissão é responsável por registrar, em banco de dados, cada etapa relevante do processo, com o objetivo de garantir rastreabilidade e de fornecer uma trilha completa de auditoria. Esses registros incluem o momento de início da submissão, os dados do pacote submetido (identificador, tipo de empacotamento, origem), às chamadas



realizadas às API do Archivematica, os status das respostas recebidas, as eventuais falhas, as reações do sistema e a confirmação da conclusão das etapas. Cada evento é gravado com as informações de data e hora, o tipo de operação e o resultado obtido e, quando aplicável, as mensagens de retorno dos serviços externos. Esse mecanismo de registro contínuo assegura a integridade da operação, permite a análise detalhada de cada submissão realizada e serve como evidência documental para auditorias internas e externas no contexto da preservação digital.

8. Após a criação do pacote **SIP** pelo Archivematica, durante o processo de ingestão, é executada uma atividade de cópia do conteúdo do SIP para uma pasta específica de armazenamento, definida no ambiente de *storage* institucional. Essa operação é realizada com o objetivo de manter uma cópia de referência do SIP antes da geração do AIP, preservando a estrutura intermediária do pacote para fins de auditoria, de reprocessamento ou de análise técnica. A cópia é executada automaticamente por um *script* ou *job* configurado para ser acionado quando o SIP atinge um estado considerado completo dentro da cadeia de microsserviços do Archivematica. O local de destino do **SIP** no *storage* é organizado de acordo com cada cenário de preservação, considerando por exemplo a volumetria dos pacotes, buscando garantir a rastreabilidade e a integridade dos pacotes armazenados. No contexto do Archivematica, o **SIP** é tratado como uma **estrutura temporária de processamento**, cujo propósito principal é viabilizar a transição do material submetido (Pré-SIP) para pacotes de preservação (AIP) e de disseminação (DIP), conforme preconiza o modelo OAIS. No entanto, ao final desse processo, o SIP **não é mantido de forma permanente no sistema**. A menos que ocorra alguma falha no processamento, ele é **descartado automaticamente** após a geração bem-sucedida dos pacotes AIP e DIP. Essa abordagem está alinhada ao princípio de que o **AIP é o pacote arquivístico de preservação definitiva**, enquanto o SIP representa apenas uma etapa intermediária, transitória e operacional. Por esse motivo, caso haja a necessidade institucional de manter o SIP como cópia de referência ou de evidência do fluxo de ingestão, é recomendável implementar um mecanismo de **cópia ou arquivamento externo antes que o Archivematica o descarte automaticamente**.
9. Ao final do processamento da ingestão, o Archivematica consolida os resultados em um pacote de preservação, **AIP** que contém o objeto digital normalizado e o conjunto de metadados, bem como os *logs* e os relatórios gerados durante o processamento. Esse **AIP** é armazenado de forma estruturada e segura no *storage service* do Archivematica. O armazenamento considera critérios de organização por identificadores únicos [*Universally Unique Identifier* (UUID)], estrutura de diretórios e integridade dos dados, assegurando que o pacote esteja disponível para futuras ações de reuso, de auditoria, de reingestão ou de acesso controlado, conforme previsto no modelo OAIS.



10. Concluído o processamento da ingestão, o Archivematica gera o **DIP**, contendo uma versão derivada do objeto digital e o conjunto de metadados necessários para fins de acesso e difusão. Esse DIP é disponibilizado em uma área específica do ambiente de *storage*, previamente configurada como ponto de integração com o sistema de disseminação. Essa área, definida como “*DIP upload target*” no Archivematica, é monitorada para que os pacotes gerados possam ser automaticamente transferidos e importados pelo sistema, conforme os parâmetros de interoperabilidade estabelecidos entre as duas aplicações. A disponibilização do DIP nessa localização estratégica garante que os conteúdos processados estejam prontos para serem incorporados ao ambiente de acesso público, assegurando a continuidade do fluxo de preservação e a efetiva disseminação dos documentos digitais.
11. Durante o processamento dos pacotes de informação, o Archivematica realiza o armazenamento sistemático de dados em seu banco de dados relacional (MySQL). Esse banco é utilizado para registrar e gerenciar informações estruturadas referentes a cada etapa do fluxo de ingestão, incluindo o status das transferências, os eventos executados pelos microsserviços, os identificadores únicos (UUID) dos pacotes, os caminhos físicos dos arquivos, os *logs* operacionais e o conjunto de metadados associados a cada objeto digital processado. Esses registros possibilitam o controle transacional do sistema, o acompanhamento do progresso das submissões e a rastreabilidade completa das ações executadas sobre cada pacote. O armazenamento no MySQL é, portanto, essencial para o gerenciamento interno da plataforma e para garantir a integridade, a consistência e a auditabilidade dos fluxos de preservação digital conduzidos pelo Archivematica.
12. Durante o fluxo de ingestão e preservação, o Archivematica utiliza o *Elasticsearch* como mecanismo de indexação e de busca para facilitar a recuperação e o monitoramento de informações operacionais. Essa indexação inclui dados sobre pacotes processados (SIP, AIP, DIP), eventos executados pelos microsserviços, identificadores (UUID), status das transferências, datas, formatos de arquivos, nomes de objetos digitais e outros metadados associados. A cada etapa relevante do processamento, o Archivematica envia documentos no formato JSON para o *Elasticsearch*, que os armazena e disponibiliza por meio de consultas rápidas e escaláveis. Essa funcionalidade é fundamental para suportar a interface administrativa do sistema, fornecer relatórios de status, permitir auditorias e oferecer recursos de busca avançada. A indexação no *Elasticsearch* contribui, assim, para a transparência e a eficiência da gestão dos processos de preservação digital.
13. O **Agente de submissão** acompanha a conclusão do fluxo de ingestão no Archivematica por meio de consultas regulares à API ou com base em notificações internas configuradas no sistema. Esse monitoramento visa identificar o momento em que o pacote de disseminação (DIP) é efetivamente gerado e armazenado na localização apropriada



(**Storage Pacotes DIP**). Uma vez detectada a finalização da ingestão e a disponibilidade do DIP, o agente ativa a rotina de integração com o sistema de disseminação, seja por meio do conector nativo do Archivematica, chamada explícita à API ou algum outro recurso requerido pelo sistema de disseminação. Essa integração tem como finalidade garantir que o conjunto de metadados e os objetos digitais resultantes da ingestão estejam devidamente registrados, organizados e publicados no sistema, tornando-os acessíveis à consulta pública por meio da interface de acesso à memória institucional.

14. Durante a ingestão dos pacotes DIP provenientes do Archivematica, o sistema de disseminação deve:
 - 14.1. Recuperar o pacote DIP no **Storage de pacotes DIP** previamente configurado como destino de disseminação no ambiente do Archivematica.
 - 14.2. Registrar, em uma fonte de dados, as informações descritivas e a vinculação dos objetos digitais. Além disso, armazena informações administrativas, como o *status* de publicação, os identificadores únicos e as relações hierárquicas entre descrições. A fonte de dados do sistema de disseminação é responsável por sustentar a navegação, a recuperação e a visualização pública dos registros, sendo essencial para garantir a integridade, a consistência e a disponibilidade das informações descritas e publicadas no sistema de disseminação.
 - 14.3. Acessar um recurso para fazer a indexação. Essa atividade garante que as derivadas de acesso dos objetos digitais preservados estejam corretamente organizadas e acessíveis para disseminação, conforme os critérios arquivísticos e os parâmetros de acesso definidos no sistema.

2.2.2 Etapas da Segunda Fase de Implementação da Arquitetura Computacional – Gestão da Preservação Digital

1. Visando oferecer suporte operacional ao gestor de preservação digital, será implementado um módulo denominado "**Relatórios e Inventários**", acessível por meio de uma interface dedicada. Essa ferramenta permitirá que o usuário, sob demanda, solicite a geração de relatórios e inventários contendo informações detalhadas sobre os pacotes de submissão (SIP) processados no ambiente de preservação digital. A interface permitirá a seleção de critérios como período, tipo de empacotamento, status da submissão ou identificadores específicos. O inventário gerado será apresentado ao usuário em formato tabular e poderá ser exportado em formatos como CSV, JSON ou PDF. Essa funcionalidade permitirá uma visão clara e rastreável dos pacotes submetidos, fortalecendo a



governança da preservação digital institucional. O sistema coletará os dados diretamente no banco de dados *SQL Server*, consolidando as informações sobre os objetos digitais ingestados, os metadados associados, a estrutura dos pacotes, as datas de submissão e os resultados do processamento. Os documentos gerados atuarão não apenas como registros formais das submissões realizadas, mas também como evidências analíticas que subsidiarão a identificação de padrões, de gargalos e de oportunidades de melhoria contínua nos fluxos de preservação digital, como também como evidência com vistas à certificação externa.

2. Com o objetivo de facilitar o acompanhamento e a tomada de decisão por parte do gestor de preservação digital, será implementado um módulo denominado ***“Dashboard para gestão”***. Essa interface gráfica fornecerá uma visão consolidada e em tempo real das atividades relacionadas aos pacotes submetidos ao sistema de preservação, incluindo dados operacionais e indicadores estratégicos.

O *dashboard* será alimentado automaticamente a partir de dados extraídos do banco de dados *SQL Server* ou de fontes indexadas, como o *Elasticsearch*, permitindo:

- Visualização do volume de submissões por período;
- Acompanhamento do status dos pacotes (em processamento, concluído, com erro etc.);
- Distribuição por tipo de empacotamento;
- Identificação dos formatos de objetos digitais mais frequentemente submetidos;
- Monitoramento da duração média dos processos de ingestão;
- Detecção de falhas ou gargalos operacionais etc.

Além de apresentar gráficos, tabelas e filtros interativos, o *dashboard* oferecerá funcionalidades para exportação de relatórios e inventários e geração de evidências que subsidiem auditorias (internas ou externas), diagnósticos e melhorias nos fluxos de preservação digital. Essa solução atuará como um componente central de transparência, rastreabilidade e governança informacional, promovendo uma gestão mais eficiente e proativa das ações de preservação institucional.

- 2.1. O sistema realizará consultas diretas ao banco de dados *SQL Server* para obter as informações necessárias à alimentação do ***“Dashboard para gestão”***, incluindo dados sobre os pacotes processados, seus status, os tipos de empacotamento, as datas de submissão e os resultados do fluxo de ingestão.



- 2.2. O sistema realizará consultas ao **Elasticsearch** para recuperar dados indexados sobre os pacotes de submissão, como eventos de processamento, status atual, tempo de execução e mensagens de *log*, complementando as informações exibidas no “**Dashboard para gestão**”.
3. Com o objetivo de apoiar a gestão estratégica, o controle institucional e a prestação de contas das ações de preservação digital, será implementado um módulo denominado “**Relatórios e Inventários de preservação**”. Essa funcionalidade permitirá a emissão de relatórios personalizados e periódicos contendo informações consolidadas sobre os fluxos de ingestão, a integridade dos objetos digitais preservados, os metadados registrados, a ocorrência de falhas e o desempenho do ambiente técnico. O gerador será acessível por meio de uma interface amigável, na qual o usuário poderá definir critérios como período, tipo de pacote (AIP, DIP), status das submissões ou faixas de duração dos processos. Com base nesses filtros, o sistema reunirá dados provenientes das diferentes bases de dados (banco relacional (MySQL), *Elasticsearch*), estruturando os relatórios em formatos compatíveis com auditorias e análise institucional (exemplo: PDF, CSV, Excel). Esses relatórios servirão como instrumentos de transparência, rastreabilidade e suporte à governança da preservação digital, podendo ser utilizados em inspeções internas, revisões de políticas arquivísticas, relatórios anuais ou apresentações institucionais.
- 3.1. O **Gerador de Relatórios e Inventários de Preservação** realizará consultas ao **banco de dados relacional (MySQL)** para obter informações estruturadas sobre os pacotes de preservação (AIP e DIP).
4. Com o objetivo de fornecer uma visão estratégica e consolidada das ações de preservação digital, será implementado o “**Dashboard para gestão de Preservação**”. Esse módulo será voltado a gestores institucionais e tomadores de decisão, apresentando indicadores-chave de desempenho, gráficos analíticos e métricas resumidas que refletirão o comportamento, a eficiência e a evolução dos fluxos de preservação digital ao longo do tempo. O *dashboard* permitirá o acompanhamento de aspectos como:
- Volume total de pacotes preservados (AIPs) por período;
 - Quantidade de objetos digitais submetidos e publicados;
 - Distribuição por tipos documentais ou áreas produtoras;
 - Taxas de sucesso e falhas nos processos de ingestão;
 - Tempos médios de processamento;



- Crescimento do acervo preservado;
- Espaço ocupado em *storage*.

Os dados serão extraídos periodicamente de fontes confiáveis, como o banco de dados relacional do Archivematica (MySQL) e o *Elasticsearch*, garantindo consistência, rastreabilidade e suporte à tomada de decisão. A interface será interativa, com filtros por período, tipo de pacote, setor produtor e outros parâmetros relevantes para análise institucional. Essa solução contribuirá para monitorar resultados, identificar tendências e planejar melhorias, promovendo uma gestão eficiente e fundamentada da preservação digital no contexto organizacional.

- 4.1. O sistema realizará consultas ao **banco de dados** para coletar indicadores gerenciais, como o número total de pacotes AIP preservados, a quantidade de objetos digitais processados, os tipos de empacotamento utilizados, as áreas produtoras envolvidas e as datas das submissões. Esses dados estruturados serão consolidados para compor os gráficos e os painéis analíticos do *dashboard*.
- 4.2. O sistema consultará o **Elasticsearch** para recuperar métricas operacionais indexadas, como o tempo de execução dos processos de ingestão, os eventos de sucesso ou de falha, as mensagens técnicas e a frequência de ocorrência de erros. Essas informações complementam a visão gerencial com dados de desempenho e confiabilidade do fluxo de preservação.

2.2.3 Auditoria da SUBMISSÃO

Concluído o processo de submissão e de disseminação do objeto digital, tanto o “**Gerador de Pré-SIP**” quanto o “**Agente de submissão**” passarão a dispor de registros completos e detalhados em suas respectivas trilhas de auditoria. Essas trilhas documentarão, cada ação executada, incluindo os identificadores atribuídos (UUID), os carimbos de data/hora, as operações realizadas, as respostas de sistemas externos, as mensagens de erro e os resultados do processamento.

Com o objetivo de garantir a transparência, a confiabilidade e a autenticidade não apenas do objeto arquivístico preservado, mas também das atividades arquivísticas e técnicas associadas à sua preservação, a solução prevê a geração de um pacote complementar de preservação contendo a trilha de auditoria da submissão anteriormente executada.

Esse pacote será estruturado de acordo com práticas consolidadas de preservação digital, podendo seguir padrões como *Metadata Encoding and Transmission Standard* (METS),



Preservation Metadata: Implementation Strategies (PREMIS) ou mesmo ser empacotado conforme os mesmos formatos utilizados nos pacotes principais. Seu conteúdo incluirá:

- Arquivo(s) contendo a trilha de auditoria do Gerador de Pré-SIP;
- Arquivo(s) contendo a trilha de auditoria do Agente de submissão;
- Metadados técnicos e administrativos sobre o processo de submissão;
- Identificadores cruzados com os pacotes Pré-SIP, SIP, AIP e, se necessário, também o DIP.

Ainda que tanto o Archivematica quanto o sistema de disseminação mantenham registros internos em seus respectivos bancos de dados relacionais — o primeiro voltado ao controle das etapas do fluxo de ingestão e de geração dos pacotes AIP e DIP, e o segundo à gestão dos registros descritivos e dos objetos digitais disseminados — essas soluções operam de forma isolada e registram apenas os eventos sob sua responsabilidade direta.

Portanto, **nenhum desses sistemas é capaz de oferecer, de forma nativa, uma visão integrada de ponta a ponta**, isto é, desde a solicitação inicial de empacotamento do objeto digital até sua efetiva disponibilização pública. Essa lacuna operacional justifica a necessidade de um componente intermediário com capacidade de orquestrar, registrar e correlacionar todos os eventos do fluxo completo de preservação digital.

Nesse contexto, o Agente de Submissão, **desenvolvido como solução intermediadora**, assume o papel central de **ponto de convergência das evidências**, sendo o componente capaz de acompanhar integralmente cada submissão — desde o momento em que o usuário solicita o empacotamento do objeto digital e seus dados, passando pela submissão ao Archivematica, o monitoramento do processamento e a geração dos pacotes SIP, AIP e DIP, até a integração e disseminação final.

Ao registrar sistematicamente cada ação executada, bem como as respostas obtidas das APIs dos sistemas envolvidos, o Agente de Submissão **garante a rastreabilidade, a auditabilidade, a autenticidade e a confiabilidade de todo o processo**, permitindo o acompanhamento operacional em tempo real, além da geração de informações para relatórios, inventários e *dashboard*.

No modelo proposto, o controle da rastreabilidade dos pacotes de submissão é iniciado no momento da geração do pacote de pré-ingestão (pré-SIP). Nessa etapa, o **“Gerador de Pré-SIP”** é responsável por gerar um identificador único (UUID), que será utilizado como referência primária para todo o fluxo de preservação digital daquele objeto. Esse UUID é inserido nos metadados do pacote, garantindo sua identificação ao longo das etapas seguintes.



Ao receber o UUID gerado pelo Gerador de Pré-SIP, o **“Agente de submissão”** inicia sua atuação no fluxo automatizado de preservação digital. Com base nesse identificador, o agente realiza uma consulta à fonte de dados, onde estão previamente armazenadas as informações referentes ao pacote gerado, incluindo o caminho de armazenamento do Pré-SIP na área de transferência, o tipo de empacotamento, os metadados associados ao processo, as credenciais de submissão e os demais parâmetros necessários.

Essa recuperação de dados permite ao agente **compor dinamicamente a requisição à API do Archivematica**, iniciando a atividade de transferência com os dados exatos do pacote que será submetido. A utilização do UUID como chave de recuperação assegura a unicidade do processo e evita inconsistências entre os componentes da solução.

Dessa forma, o Agente de Submissão atua de maneira autônoma e precisa, mantendo a coesão entre os dados estruturados no banco, o conteúdo físico do pacote de pré-ingestão e a lógica de submissão via API, fortalecendo a confiabilidade do processo de preservação digital e garantindo a rastreabilidade completa, desde a origem do empacotamento até a ingestão no ambiente arquivístico.

Durante a execução do fluxo de preservação digital, o Agente de Submissão, **desenvolvido como parte da solução de automação da ingestão de objetos digitais**, desempenha um papel central na comunicação com a API do Archivematica. Um dos pontos críticos desse processo é a obtenção e o gerenciamento do identificador UUID, que é atribuído pelo Archivematica a cada submissão de pacote de informação.

No momento em que o **“Agente de Submissão”** executa a chamada ao *endpoint* POST `/api/transfer/start_transfer/`, fornecendo os parâmetros do pacote (nome, tipo, caminho e configurações de acesso), o Archivematica processa a solicitação e retorna uma resposta contendo, entre outros elementos, o UUID exclusivo da submissão iniciada. Esse identificador é retornado no campo **id** da resposta JSON e representa a chave de rastreamento da transferência ao longo de todo o fluxo de ingestão.

A captura imediata desse UUID pelo agente é essencial para garantir a rastreabilidade da submissão. O identificador é armazenado em uma fonte de dados, juntamente com os metadados do processo (por exemplo, número do objeto digital, tipo de empacotamento e data da submissão), permitindo que etapas subsequentes — como a aprovação manual da transferência, o monitoramento do progresso da ingestão, a identificação do AIP resultante e a integração com o sistema de disseminação — sejam realizadas com precisão e consistência.

Além disso, o UUID é utilizado como elemento-chave nos mecanismos de auditoria e de geração de inventários, viabilizando a associação entre o pacote digital submetido e sua representação no repositório arquivístico. Portanto, a obtenção e o correto gerenciamento



do UUID pelo Agente de Submissão constituem um componente fundamental da arquitetura de preservação digital proposta, assegurando o controle, a autenticidade, confiabilidade e a reprodutibilidade dos processos realizados de forma automatizada.

Com vistas a assegurar a rastreabilidade integral do fluxo de preservação digital, o Agente de Submissão persiste e gera **dois identificadores distintos para cada pacote**:

1. **UUID-Empacotador** – gerado no instante em que o módulo monta o pré-SIP;
2. **UUID-Archivematica** – retornado pelo *endpoint* `/api/transfer/start_transfer/` ao aceitar a transferência.

Logo após receber o UUID-Empacotador, o agente consulta a base relacional, recupera os metadados do pacote e efetua a chamada à API do Archivematica. A resposta dessa chamada contém o UUID-Archivematica, que o agente captura e grava na mesma linha de registro, formando uma **chave composta (UUID-Empacotador ↔ UUID-Archivematica)**.

Essa associação é crítica por três razões:

- **Rastreamento de ponta a ponta** – permite correlacionar, sem ambiguidade, a solicitação de empacotamento, o processamento no Archivematica e a disseminação em uma plataforma de acesso, difusão ou transparência ativa.
- **Auditoria e recuperação** – qualquer evento ou anomalia registrado nos sistemas envolvidos pode ser localizado, usando um ou os dois UUID, para fins de auditoria/recuperação.
- **Governança de objetos digitais** – o par de identificadores fundamenta relatórios e inventários gerenciais, *dashboards* e análises de desempenho, pois une dados que, de outra forma, permaneceriam fragmentados entre aplicações.

2.2.4 Sequência de passos da AUDITORIA DA SUBMISSÃO

1. Finalizada a submissão ao Archivematica, o “**Agente de Submissão**” aciona o módulo “**Gerador de Pré-SIP**” para iniciar a consolidação da trilha de auditoria da submissão. Esse acionamento é feito por meio de chamada agendada, via *job* no **Hangfire**), informando o UUID da submissão e os parâmetros necessários para a coleta dos registros de auditoria.
2. O “**Gerador de Pré-SIP**”, após ser acionado pelo Agente de Submissão, inicia a coleta dos registros de auditoria armazenados em banco de dados (SQL Server). Com base nesses dados, gera um pacote no formato Pré-SIP.



3. Após sua geração, o pacote Pré-SIP é armazenado em *storage* (**Storage Pacotes Pré-SIP**).
4. Após a finalização da geração do pacote Pré-SIP de auditoria, o "**Gerador de Pré-SIP**" aciona o "**Agente de submissão**", informando o UUID do processo de auditoria recém-criado. Com esse acionamento, o controle da operação é transferido para o agente, que dará início à submissão automatizada do pacote de auditoria para o ambiente do Archivematica, utilizando as mesmas rotinas seguras e rastreáveis aplicadas ao envio de pacotes. Esse acionamento permite manter o processo padronizado, auditável e tecnicamente alinhado com o fluxo institucional de preservação.
5. O "**Agente de submissão**", ao ser acionado, inicia o fluxo de transferência do pacote de auditoria no Archivematica por meio de uma chamada à **API de Transfer** (/api/transfer/start_transfer/). Nessa solicitação, são informados os dados essenciais, como o nome do pacote, o tipo de transferência, o caminho relativo no *storage* e o UUID do pacote de auditoria gerado. A **API de Transfer** do Archivematica obtém o Pacote Pré-SIP do Storage (**Storage Pacotes Pré-SIP**) e inicia automaticamente a cadeia de microsserviços responsável por validar, identificar e preparar o pacote para a próxima etapa de ingestão, conforme os parâmetros definidos no *pipeline* configurado.
6. O "**Agente de submissão**" aciona o Archivematica para iniciar a atividade de ingestão do pacote de auditoria por meio de chamada de **API de Ingest**.
7. Durante o fluxo de ingestão do pacote de auditoria, o Archivematica executa a geração do SIP. Concluída essa etapa, o conteúdo do SIP é automaticamente copiado para uma área específica do *storage* institucional (**Storage Pacotes SIP**), definida previamente para esse fim. Essa cópia tem como objetivo preservar a versão intermediária do pacote, garantindo rastreabilidade e possibilitando auditorias técnicas futuras.
8. Ao final do processamento do pacote de auditoria, o Archivematica gera o AIP, que consolida o conteúdo do pacote SIP, os metadados de auditoria estruturados e os *logs* das etapas realizadas. Esse AIP representa a evidência formal e autenticada do processo de submissão, incluindo tanto os registros operacionais do "**Gerador de Pré-SIP**" quanto do "**Agente de submissão**". Em seguida, o AIP é armazenado no *Storage service* (**Storage Pacotes AIP**).
9. Após a geração e o armazenamento do AIP de auditoria, caso o cenário de preservação requeira a criação de um DIP correspondente, ele será gerado de forma automática. Esse DIP contém uma versão derivada e acessível da trilha de auditoria, incluindo o conjunto mínimo de metadados e relatórios estruturados sobre o fluxo de empacotamento e de submissão executados. O pacote é, então, disponibilizado em uma área previamente



configurada como destino de disseminação no ambiente de *storage* (**Storage Pacotes DIP**), destinada à integração com sistema de disseminação.

10. Uma vez detectada a finalização da ingestão e a disponibilidade do DIP (caso necessário), o “**Agente de Submissão**” poderá ativar a rotina de integração com o sistema de disseminação, seja por meio do conector nativo do Archivematica ou por uma chamada explícita à API ou à ferramenta de importação do sistema de disseminação.

Com a realização da disseminação do pacote preservado e a incorporação do pacote de auditoria no sistema de disseminação, conclui-se não apenas o ciclo técnico de preservação digital, mas também a promoção da transparência e da rastreabilidade de todas as etapas envolvidas. Ao garantir que tanto o objeto arquivístico quanto os registros das ações de preservação sejam descritos, armazenados e acessíveis, a instituição fortalece sua governança informacional, assegura a conformidade com boas práticas arquivísticas e oferece suporte à *accountability* institucional.