

**Projeto de Pesquisa Acervo Digital  
Semântico Da Funarte: aplicação de  
novos modelos de organização e  
interconexão da informação dos  
acervos digitalizados do CEDOC**

**Relatório referente à Meta 2 do TED  
Ibict e Funarte**

**Metadados dos acervos digitalizados  
tratados e normalizados**

**Março/2021**

## Sumário

1. Introdução.....	2
2. Diagnóstico interconexão dos softwares dos repositórios digitais ..	3
2.1. O OAI-PMH como protocolo de comunicação entre os repositórios..	4
2.2. Diagnóstico do Sistema SophiA Biblioteca (Catálogo CEDOC) .....	6
2.3. Diagnóstico do sistema AtoM (Arquivos Privados).....	8
2.4. Diagnóstico do sistema SophiA Acervo (Sérgio Britto Digital).....	10
3. Implementação do processo de interconexão dos repositórios ....	11
4. Resultados e encaminhamentos.....	14
Referências .....	16

## 1. Introdução

Este documento é referente ao segundo produto do projeto intitulado “Acervo Digital Semântico da Funarte: Aplicação de Novos Modelos de Organização e Interconexão da Informação dos Acervos Digitalizados do CEDOC”, realizado a partir da parceria entre a Fundação Nacional de Artes (Funarte) e o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

O projeto tem como objetivo principal reformular a disponibilização dos acervos do centro de documentação e pesquisa da Funarte (CEDOC), em busca de aplicar um modelo de organização e interconexão dos acervos do departamento disponíveis no sítio da instituição<sup>1</sup>. O segundo produto aqui tratado, por sua vez, dá sequência aos procedimentos realizados no primeiro produto (será mencionado no próximo parágrafo), e descreve o processo de tratamento e normalização dos metadados dos acervos do CEDOC mapeados para a organização e interconexão.

Para contextualizar melhor a sequência das atividades que serão apresentadas, vale recuperar os encaminhamentos do primeiro produto do projeto, denominado “Modelo conceitual de interoperabilidade semântica dos acervos”. Ao todo, foram executadas 3 atividades principais no primeiro produto:

- Apresentação da proposta do modelo conceitual de interoperabilidade semântica;
- Avaliação de mudanças na página do CEDOC;
- Apresentação uma proposta de reformulação da página do CEDOC.

Como fruto do primeiro produto então, foram diagnosticados três repositórios digitais diferentes pertencentes ao CEDOC: o acervo do *Catálogo CEDOC*, o acervo dos *Arquivos Privados*, e o acervo *Sérgio Brito Digital*. Para implementar a interconexão entre os três repositórios foi proposto um modelo de interoperabilidade entre eles, utilizando a plataforma aberta Tainacan como repositório integrador, e o protocolo de comunicação OAI-PMH para troca de metadados entre os repositórios.

---

<sup>1</sup> Sítio da Funarte - <https://www.funarte.gov.br/>

Assim, neste relatório serão apresentados os avanços em busca de aplicar o modelo de interconexão proposto nos produtos anteriores, indicando os resultados do diagnóstico realizado sobre os softwares dos repositórios digitais, bem como apresentados os caminhos de interconexão tomados de acordo com as possibilidades e implementação do modelo proposto.

Para isso, serão tratados nos próximos tópicos deste relatório:

- **Diagnóstico de interconexão dos softwares dos repositórios digitais:** serão apresentadas as análises das possibilidades de conexão dos dados dos repositórios digitais do CEDOC com o Tainacan.
- **Implementação do processo de interconexão dos repositórios:** serão apresentadas as atividades realizadas em busca de efetivar a interconexão dos repositórios, de acordo com o diagnóstico realizado previamente.
- **Resultados e encaminhamentos:** serão retomados os resultados desse segundo produto, bem como os encaminhamentos para o próximo produto do projeto.

## 2. Diagnóstico de interconexão dos softwares dos repositórios digitais

O objetivo do processo de diagnóstico da interconexão dos softwares dos repositórios digitais do CEDOC foi entender quais seriam os requisitos para colocar em prática a agregação desses repositórios. Sendo assim, nos próximos parágrafos serão descritos os estudos dos protocolos de comunicação de cada um dos três sistemas, bem como os encaminhamentos técnicos realizados para a implementação da agregação dos acervos.

Primeiramente, é importante esclarecer o que são esses protocolos de comunicação. No contexto deste produto, protocolo de comunicação é a funcionalidade do sistema que permite que os dados do repositório sejam disponibilizados para coleta por outro sistema, que deve possuir o mesmo protocolo de comunicação, permitindo assim, efetuar uma ponte entre os sistemas, por onde os dados serão compartilhados.

O modelo de interconexão proposto no primeiro produto deste projeto leva em conta um diagnóstico prévio, que apresentou que os sistemas SophiA

Biblioteca, AtoM e Tainacan possuem o mesmo protocolo de comunicação, o OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) e, por isso, podemos pensar uma agregação dos acervos utilizando o Tainacan no contexto de repositórios da Funarte.

## 2.1. O OAI-PMH como protocolo de comunicação entre os repositórios

O OAI-PMH é um protocolo de comunicação entre sistemas, e tem duas funções principais: disponibilizar metadados, e fornecer um serviço de agregação. Na prática, esse protocolo existe como uma funcionalidade do repositório digital, que permite disponibilizar dados para coleta e, ao mesmo tempo, obter dados de outros sistemas que também utilizem o OAI-PMH.

Como mostra a Figura 1 abaixo, o contexto de agregação de repositórios através de um protocolo de comunicação do tipo OAI é composto de dois agentes: os **provedores de dados**, de onde os dados serão obtidos para agregação (no caso deste projeto, são considerados provedores de dados o Catálogo CEDOC, o Acervo Sérgio Britto Digital e os Arquivos Privados); e o **provedor de serviço**, que é o repositório no qual os dados serão agregados, organizados e estarão disponíveis para a consulta (no caso atual, seria o papel do Tainacan).

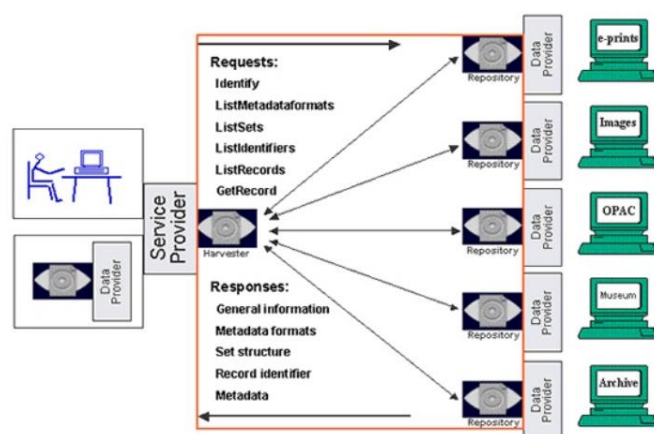


Figura 1 - Esquema de comunicação através de protocolos do tipo OAI. (IBICT, 2021)

Ainda como mostra a Figura 1 acima, para efetivar a comunicação entre os repositórios, o provedor de serviços faz consultas aos provedores de dados, e estes retornam como respostas seus dados estruturados. Essa comunicação é efetivada através de uma URL específica que indica qual é o ponto de acesso

ao OAI-PMH dos repositórios, uma vez identificada essa URL, as consultas são realizadas através de requisições específicas e, assim, os dados são obtidos como resposta às requisições, possibilitando por fim sua agregação em um provedor de serviço.

As requisições existentes são (OPEN AERCHIVES, 2021):

- GetRecord: que é utilizada para obter um registro em específico do provedor de dados, para isso é preciso ter o número identificador desse registro, bem como o padrão de metadados para o qual ele foi mapeado no contexto do OAI-PMH.
- Identify: é utilizada para obter informações que identifiquem o repositório.
- ListIdentifiers: é utilizada para obter as informações básicas dos registros do repositório, bem como seus números de identificação. É obrigatório informar qual o padrão de metadados, e se houver, a chave de acesso ao repositório OAI-PMH.
- ListMetadataFormats: é utilizada para obter as informações sobre em quais padrões de metadados os registros estão disponíveis.
- ListRecords: é utilizada para fazer a coleta dos dados completos dos registros do repositório. É obrigatório informar qual o padrão de metadados, e se houver, a chave de acesso ao repositório OAI-PMH.
- ListSets: é utilizada para obter a estrutura do repositório, como por exemplo, as coleções existentes. É obrigatório informar, se houver, a chave de acesso ao repositório OAI-PMH.

Como pode-se observar nas requisições disponíveis, duas informações principais são necessárias para a coleta dos dados:

- Padrão de Metadados: para que um repositório esteja efetivamente disponível através do OAI-PMH, é necessário que seus registros estejam mapeados para um padrão de metadados, que comumente é o Dublin Core.
- Chave de Acesso: só é necessária se houver alguma restrição de acesso aos dados dos provedores, como estamos trabalhando

somente com dados liberados ao público, não deve ser necessário restringir o acesso aos dados.

Assim, com esses fatores estabelecidos: a **URL de acesso ao OAI-PMH**, e devidamente mapeado o **padrão de metadados** e preferencialmente **sem restrição de acesso**, será possível configurar a conexão entre os repositórios provedores de dados e o repositório provedor de serviço.

Nós próximos tópicos serão descritos os procedimentos realizados para identificar esses fatores necessários para efetivar a conexão via OAI-PMH nos repositórios do CEDOC.

## 2.2. Diagnóstico do Sistema SophiA Biblioteca (Catálogo CEDOC)

O SophiA Biblioteca<sup>2</sup> é um software para gestão de bibliotecas fornecido pela Prima<sup>3</sup>, e foi a solução adotada pela Funarte para gerenciar os acervos das bibliotecas da Funarte, denominado Catálogo CEDOC:

“O CEDOC é composto basicamente dos acervos das bibliotecas de três fundações de caráter nacional extintas em 1990: a Fundação Nacional de Arte/Funarte, a Fundação Nacional de Artes Cênicas/Fundacen e a Fundação do Cinema Brasileiro/FCB. Essas três bibliotecas foram absorvidas pela atual Funarte, reunindo volumosa documentação (cerca de 1 milhão de itens) sobre Artes Plásticas e Gráficas, Música, Fotografia, Teatro, Dança, Ópera, Circo, Cinema e Vídeo.” (FUNARTE, 2021)

Então, o acervo bibliográfico da Funarte é constituído ao longo da existência da instituição, cujo conteúdo digitalizado e/ou catalogado está disponível para consulta on-line através da interface de busca do SophiA Biblioteca<sup>4</sup>, estima-se em torno de 125 mil registros disponíveis para consulta através do sistema.

Relembrando uma constatação obtida no produto anterior, o SophiA Biblioteca possui a funcionalidade OAI-PMH, como consta no site *Portal Sophia* (2021),

“Moderno e de fácil utilização, o software baseia-se nos padrões internacionais de catalogação e comunicação de dados (MARC-21, ISO2709, Z39.50 cliente e servidor, XML e OAI-PMH), e contabiliza mais de 600 instituições usuárias, entre universidades, escolas, empresas, órgãos culturais e públicos, no Brasil e exterior.”

<sup>2</sup> Sobre o SophiA Biblioteca - <http://www.portalsophia.com.br/SobreBiblioteca.aspx>

<sup>3</sup> Prima - <http://www.portalsophia.com.br/SobrePrima.aspx>

<sup>4</sup> Catálogo CEDOC - [http://cedoc.funarte.gov.br/sophia\\_web/index.html](http://cedoc.funarte.gov.br/sophia_web/index.html)

É a partir da existência do OAI-PMH no sistema que será possível realizar a coleta dos dados do acervo e prosseguir com a agregação com os demais acervos.

Dessa forma foi enviada a solicitação da URL de acesso do OAI-PMH do Sophia Web à equipe do CEDOC, que como mencionado nos parágrafos finais da introdução, é necessário para dar início à configuração da coleta dos dados.

Em um primeiro momento, a equipe do CEDOC não tinha conhecimento de um ponto de acesso ao OAI-PMH do sistema, sendo necessário estabelecer contato com o suporte da Prima. A equipe do CEDOC então identificou junto à Prima que a funcionalidade do OAI-PMH é tratada como um módulo do sistema, que precisava ser instalado para que a URL de acesso fosse disponibilizada.

A URL disponibilizada<sup>5</sup> para acessar o OIA-PMH do Catálogo CEDOC funciona, e ainda permite fazer as requisições normalmente, como apresenta o exemplo da Figura 2 abaixo, que mostra o resultado da requisição *Identify*:

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2021-03-01T04:10:36Z</responseDate>
  <request verb="Identify">http://cedoc.funarte.gov.br/sophia_web/oai/Imagens_e_Arquivos/oai.asp</request>
  <Identify>
    <repositoryName>Imagens_e_Arquivos</repositoryName>
    <baseURL>http://cedoc.funarte.gov.br/sophia_web/oai/Imagens_e_Arquivos/oai.asp</baseURL>
    <protocolVersion>2.0</protocolVersion>
    <adminEmail/>
    <earliestDatestamp>2015-09-29T09:54:25Z</earliestDatestamp>
    <deletedRecord>no</deletedRecord>
    <granularity>YYYY-MM-DDThh:mm:ssZ</granularity>
    <description>
      <?xml version="1.0" encoding="UTF-8" ?>
      <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
        <scheme>oai</scheme>
        <repositoryIdentifier/>
        <delimiter></delimiter>
        <sampleIdentifier>oai::1002</sampleIdentifier>
        </OAI-PMH>
      </description>
    </Identify>
  </OAI-PMH>
```

Figura 2 - Resposta da requisição *Identify*<sup>6</sup> do OAI-PMH do Catálogo CEDOC.

Porém, não houve padrões de metadados identificados na resposta à requisição *ListMetadataFormats*, que retornou vazia, indicando que não houve um processo de mapeamento dos metadados para um padrão aceito pelo OAI-PMH, ou que há algum problema em apresentar os padrões de metadados disponíveis, como apresenta a Figura 3 abaixo:

<sup>5</sup> URL do OAI-PMH do Catálogo CEDOC -

[http://cedoc.funarte.gov.br/sophia\\_web/oai/Imagens\\_e\\_Arquivos/oai.asp](http://cedoc.funarte.gov.br/sophia_web/oai/Imagens_e_Arquivos/oai.asp)

<sup>6</sup> Link para a requisição *Identify* -

[http://cedoc.funarte.gov.br/sophia\\_web/oai/Imagens\\_e\\_Arquivos/oai.asp?verb=Identify](http://cedoc.funarte.gov.br/sophia_web/oai/Imagens_e_Arquivos/oai.asp?verb=Identify)



This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2021-03-01T09:33:32Z</responseDate>
  <request verb="ListMetadataFormats">
    <ListMetadataFormats></ListMetadataFormats>
  </request>
</OAI-PMH>
```

Figura 3 - Resposta da requisição ListMetadataFormats do OAI-PMH do Catálogo CEDOC.

Isso significa que, sem um padrão de metadados configurado no OAI-PMH, não é possível fazer a requisição pelos dados do repositório. Como mencionado no tópico 2.1, além da URL do OAI-PMH, é necessário que exista um padrão de metadados configurado, pois é informação obrigatória na requisição dos registros, caso contrário não há como requisitar os registros.

Essa condição foi explicada à equipe do CEDOC, que tentou entender possíveis soluções com a Prima, porém não houve avanços na solução do problema, então o resultado do diagnóstico é a impossibilidade de efetivar a conexão do acervo Catálogo CEDOC na estrutura de interconexão de acervos proposta.

### 2.3. Diagnóstico do sistema AtoM (Arquivos Privados)

O AtoM (Access to Memory) é uma aplicação web gratuita indicada para descrição arquivística, é um software amplamente utilizado para construção de repositórios arquivísticos, e este é o caso de sua aplicação na Funarte. A adoção do AtoM pelo CEDOC<sup>7</sup> ainda está sendo implementada, e prevê a criação de um acervo de arquivos privados sobre artistas brasileiros, com conjuntos de dossiês compostos por imagens dos artistas, e estruturados por fundos arquivísticos.

Diferente do SophiA Biblioteca, o AtoM tem uma área de documentação própria para descrever a funcionalidade OAI-PMH no sistema<sup>8</sup>, isso facilita a identificação da URL de acesso, e ainda apresenta de maneira breve como executar requisições, e o que se deve esperar como resposta.

Dessa forma, a primeira tarefa do diagnóstico de interconexão do AtoM foi testar a URL de acesso indicada na documentação, que é formada pelo link do acervo (<https://atom.funarte.gov.br/>) acrescido de `;oai?verb=Identify`, já utilizando a requisição *identify* para apresentar as informações de identificação

<sup>7</sup> Acervo de Arquivos Privados do CEDOC - <https://atom.funarte.gov.br/>

<sup>8</sup> OAI-PMH no AtoM - <https://www.accesstomemory.org/pt-br/docs/2.6/user-manual/import-export/oai-pmh/>

do OAI-PMH. Em um primeiro momento esse resultado não retornou nada, somente uma página em branco, indicando que a funcionalidade do OAI-PMH não estava ativa, o que levou ao contato com a equipe do CEDOC, que acionou os responsáveis pelo repositório, e então a funcionalidade foi ativada e a URL de acesso<sup>9</sup> efetivamente apresentou resultados (Figura 4).

This XML file does not appear to have any style information associated with it. The document tree is shown below:

```
<?xml version="1.0" xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2021-03-01T14:12:44Z</responseDate>
  <request verb="Identify">https://atom.funarte.gov.br/;oai</request>
  <Identify>
    <repositoryName>FUNARTE</repositoryName>
    <baseURL>https://atom.funarte.gov.br/oai/index.php/</baseURL>
    <protocolVersion>2.0</protocolVersion>
    <adminEmail>oai-pmh-teste@brasil.gov.br</adminEmail>
    <lastModifiedDate>2019-05-08T07:35:34Z</lastModifiedDate>
    <deletedRecord/>
    <granularity>http://www.openarchives.org/OAI/2.0/granularity</granularity>
    <compression>gzip</compression>
  </Identify>
  <description>
    <oai-identifier xmlns="http://www.openarchives.org/OAI/2.0/oai-identifier" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai-identifier http://www.openarchives.org/OAI/2.0/oai-identifier.xsd">
      <identifier>atom.funarte.gov.br</identifier>
      <delete/>
      <sampleIdentifier>oai:atom.funarte.gov.br:COD54RQFUNARTE_100002</sampleIdentifier>
    </oai-identifier>
  </description>
</Identify>
</OAI-PMH>
```

Figura 4 - Resposta da requisição Identify do OAI-PMH dos Arquivos Privados.

Outra necessidade para implementar a conexão desse repositório como provedor de dados é a existência de um padrão de metadados mapeado, que pode ser encontrado fazendo a requisição *ListMetadataFormats*, que ao ser requisitada apresentou como resposta a existência do *oai\_dc* (Dublin Core) e o *oai\_ead* (Encoded Archival Description) como padrões de metadados mapeados para os registros do repositório (Figura 5).

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2021-03-01T14:12:44Z</responseDate>
  <request verb="ListMetadataFormats">https://atom.funarte.gov.br/;oai</request>
  <ListMetadataFormats>
    <metadataFormat>
      <metadataPrefix>oai_dc</metadataPrefix>
      <schema>http://www.openarchives.org/OAI/2.0/oai_dc.xsd</schema>
      <metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc/</metadataNamespace>
    </metadataFormat>
    <metadataFormat>
      <metadataPrefix>oai_ead</metadataPrefix>
      <schema>http://www.loc.gov/ead/ead.xsd</schema>
      <metadataNamespace>urn:isbn:1-931666-22-9</metadataNamespace>
    </metadataFormat>
  </ListMetadataFormats>
</OAI-PMH>
```

Figura 5 - Resposta da requisição ListMetadataFormats do OAI-PMH dos Arquivos Privados

Porém, mesmo com o acesso efetivo à essas requisições de identificação, o acesso aos registros estava restrito, pois o OAI-PMH havia sido configurado com uma chave de acesso, que nos foi passada. Contudo, como os registros compartilhados são públicos e não necessitam de restrição, foi sugerido

<sup>9</sup> URL de acesso ao OAI-PMH do AtoM = <https://atom.funarte.gov.br/;oai?verb=Identify>

à equipe CEDOC que não limitasse o acesso aos registros, e deixasse a requisição livre. E assim, atualmente é possível listar os registros do repositório de Arquivos Privados normalmente através do OAI-PMH (Figura 6).

This XML file does not appear to have any style information associated with it. The document tree is shown below:

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2021-03-03T14:27:43Z</responseDate>
  <request verb="ListRecords" metadataPrefix="oai_dc">https://atom.funarte.gov.br:/oai/</request>
  <ListRecords>
    <record>
      <header>
        <identifier>oai:atom.funarte.gov.br:CODEARQFUNARTE_5</identifier>
        <timestamp>2020-08-14T14:57:12Z</timestamp>
        <setSpec>oai:atom.funarte.gov.br:CODEARQFUNARTE_5</setSpec>
      </header>
      <metadata>
        <?xml version="1.0" encoding="UTF-8" ?>
        <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
          <dc:title>Coleção Aderbal Júnior</dc:title>
          <dc:creator>Aderbal Júnior</dc:creator>
          <dc:description>Doação formada, essencialmente, por questionários preenchidos pelos espectadores do espetáculo "Amanhã, Amélia, amanhã", de Leilah Assumpção, direção de Aderbal Freire-Filho, apresentado entre 04/07/1973 e 12/07/1973.</dc:description>
          <dc:format>0,12 m.</dc:format>
          <dc:identifier>https://atom.funarte.gov.br/fundo-aderbal-junior</dc:identifier>
          <dc:relation>https://atom.funarte.gov.br/centro-de-documentacao-e-informacao-cedoc-da-funarte</dc:relation>
          <dc:relation>Centro de Documentação e Informação - CEDOC - FUNARTE</dc:relation>
        </oai_dc:dc>
      </metadata>
    </record>
  </ListRecords>
</OAI-PMH>
```

Figura 6 - Resposta da requisição ListRecords do OAI-PMH dos Arquivos Privados

De todo modo, na data de execução dessa etapa do projeto houve uma reunião com a equipe do CEDOC para apresentar os resultados do diagnóstico aqui relatado, foi mencionado que o AtoM ainda está em implementação, e os dados do acervo não estão completos o suficiente para prosseguir com a agregação deste repositório.

Dessa forma, conclui-se com esse diagnóstico, que o repositório de Arquivos Privados disponibilizado pelo AtoM é passível de conexão com o provedor de serviços (Tainacan) através do OAI-PMH, porém, devido a incompletude do acervo não é recomendado efetivar a interconexão deste repositório, já que isso pode causar divergências nos dados agregados.

## 2.4. Diagnóstico do sistema SophiA Acervo (Sérgio Britto Digital)

O SophiA Acervo é um dos softwares oferecidos pela Prima, assim como o SophiA Biblioteca, porém com um foco voltado para catalogação de coleções de conteúdo multimídia. Para este software não foram encontradas informações sobre a existência do OAI-PMH, e as tentativas de identificar a URL de acesso (adicionando /oai/ - ;oai - /oai.asp? e etc.) não deram resultado.

Dessa forma, o diagnóstico sobre a conexão através do OAI-PMH foi de impossibilidade de estabelecer a interconexão deste repositório através desse protocolo. Uma alternativa proposta em conjunto com a equipe do CEDOC foi realizar a migração do acervo do SophiA Acervo para o Tainacan, entendendo

que o Tainacan, já sendo um software adotado pela instituição, tem plenas condições de manter um acervo com o conteúdo majoritariamente fotográfico do Sérgio Britto Digital, além das dificuldades que a equipe do CEDOC já havia relatado de gerenciar o acervo utilizando o sistema SophiA.

Assim, o encaminhamento final do diagnóstico desse repositório foi a exportação dos dados e imagens do acervo pela equipe do CEDOC<sup>10</sup>, e que esses dados seriam posteriormente migrados para o Tainacan pela equipe do IBICT.

### 3. Implementação do processo de interconexão dos repositórios

A proposta de interconexão de repositórios, fruto do produto 1 deste projeto, partia da premissa de que o protocolo de comunicação OAI-PMH estaria em pleno funcionamento nos sistemas geridos pelo CEDOC, como isso não aconteceu como previsto, um caminho alternativo para a interconexão dos acervos precisou ser desenvolvido.

Assim, a equipe do CEDOC resolveu manter no plano de trabalho a migração do acervo Sérgio Britto Digital para o Tainacan, e incluir a criação de uma área no site da Funarte denominada CEDOC Digital, que se manterá no contexto das páginas referentes ao CEDOC, e agregará o acervo Sérgio Brito Digital e uma parcela de objetos já digitalizados do acervo de Arquivos Privados.

A motivação da criação do CEDOC Digital partiu da necessidade da equipe do CEDOC de tornar mais aprazível a forma como o acervo é exposto aos usuários e, como o Tainacan tem funcionalidades de exibição dos objetos do acervo nas páginas do WordPress, a proposta é criar páginas sobre os objetos desses acervos que servirão de “vitrine” para os usuários, e que eles eventualmente sejam redirecionados aos repositórios de origem (Catálogo CEDOC, Arquivos Privados).

Desse modo foi realizada, ainda dentro deste produto, a migração do acervo Sérgio Britto Digital para o Tainacan. Com essa migração espera-se

---

<sup>10</sup> Como o SophiA Acervo não tem uma documentação disponível de forma acessível ao usuário, e a interface também não indica de forma efetiva como efetuar a exportação de dados e imagens do repositório, só foi possível a exportação de uma tabela de dados. A coleta de imagens teve que ser feita a partir de uma raspagem de dados da interface web do acervo, realizada pela equipe do IBICT. A descrição desse processo estará descrita no relatório 3 deste projeto.

modificar a forma como a equipe do CEDOC pode interagir com os objetos do acervo, dando oportunidade para a criação de exposições digitais através do Tainacan.

Para realizar a migração dos dados, foi solicitado inicialmente que a equipe do CEDOC fizesse a exportação dos dados e imagens, porém, o sistema SophiA Acervo se demonstrou pouco intuitivo no processo de obtenção dos dados, principalmente na obtenção das imagens, dificultando a identificação de qual o caminho era necessário percorrer para exportar o acervo.

Dessa forma, nos foi cedido um acesso remoto à interface de gestão do sistema, onde foi identificado o número total de registros do acervo (4.205), até então desconhecido, e foi possível também obter uma planilha com os dados sobre cada objeto catalogado no acervo, porém, sem sucesso na obtenção das imagens.

Para obter as imagens foi desenvolvido um *script* em *Python* de raspagem de dados<sup>11</sup>, que efetuou a coleta do *link* de cada imagem pela interface web do acervo<sup>12</sup>. O processo envolveu acessar a planilha de dados exportada do sistema, e pesquisar título por título na caixa de pesquisa do acervo on-line. Ao obter os resultados, foi necessário fazer uma comparação entre os dados apresentados na interface web e os dados da planilha exportada e, assim, ter certeza de que as imagens correspondiam a seus respectivos registros.

Uma vez realizada a coleta dos *links* das imagens para cada registro, foi desenvolvido um outro *script*<sup>13</sup> em *Python* para realizar o download de cada imagem pelo seu *link* coletado e, dessa forma, armazenar as imagens originais do acervo em disco físico.

De posse da planilha de dados e das imagens originais, foi criada uma coleção no Tainacan da Funarte, referente ao acervo Sérgio Britto Digital, em que os metadados do acervo foram replicados na coleção e a planilha de dados importada.

---

<sup>11</sup> Script para coleta de imagens -

[https://github.com/tainacan/data\\_science/blob/master/FUNARTE/coleta\\_imagens\\_SB.py](https://github.com/tainacan/data_science/blob/master/FUNARTE/coleta_imagens_SB.py)

<sup>12</sup> Interface web do acervo Sérgio Britto Digital (SophiA Acervo) -

[http://sbrittod.funarte.gov.br/sophia\\_acervo/](http://sbrittod.funarte.gov.br/sophia_acervo/)

<sup>13</sup> Script para download as imagens -

[https://github.com/tainacan/data\\_science/blob/master/FUNARTE/download\\_images\\_by\\_url.py](https://github.com/tainacan/data_science/blob/master/FUNARTE/download_images_by_url.py)

Já quanto às imagens, foi solicitado pela equipe do CEDOC uma redução do DPI (Dots Per Inch) de todas as imagens, para evitar que estas fossem impressas com a qualidade original. Para isso foi novamente desenvolvido um *script*<sup>14</sup> em *Python* que realiza a alteração do DPI de todas as imagens para 72 produzindo, assim, dois conjuntos de imagens: as imagens originais sem alteração; e as imagens para publicação no acervo com o DPI reduzido.

Mesmo assim, como as imagens só estavam disponíveis localmente, foi necessário disponibilizar essas imagens tanto para o servidor de arquivos (qualidade original), quanto o servidor web (qualidade reduzida) da Funarte, e garantir que a equipe do CEDOC tenha a posse dos dois conjuntos de imagens.

Para enviar as imagens com qualidade reduzida, foi necessário passar por um processo de homologação junto a TI da Funarte, para garantir que a importação delas no Tainacan não causasse nenhum problema ao site da instituição. Esse processo envolveu enviar as imagens para um servidor web de homologação (que representa uma cópia do servidor onde hoje está o site da Funarte), e recriar a coleção Sérgio Britto Digital no Tainacan, para assim importar as imagens para essa coleção, e uma vez validado esse processo, as imagens seriam migradas para o servidor atual do site, e finalmente poderiam ser importadas no Tainacan.

As imagens originais foram enviadas em outro processo, visto que essas devem ficar no servidor de arquivos da Funarte, disponível para consulta individual dos gestores e cobertas por uma segurança de backup e armazenamento. Para isso, novamente com o apoio da equipe de TI, foram enviadas as imagens para o servidor de arquivos através de um FTP (File Transfer Protocol), e então esse conjunto de imagens originais já está sob a gestão efetiva da equipe do CEDOC.

O resultado de todo esse processo de migração é o acervo do Sérgio Britto Digital, que está disponível ao acesso público através do Tainacan (<https://www.funarte.gov.br/acervo-sergio-britto-digital/>). O próximo passo agora é criar uma página de exibição deste acervo para integrá-la a área proposta do CEDOC Digital.

---

<sup>14</sup> Script para redução de DPI das imagens - [https://github.com/tainacan/data\\_science/blob/master/FUNARTE/reduce\\_dpi.py](https://github.com/tainacan/data_science/blob/master/FUNARTE/reduce_dpi.py)

#### 4. Resultados e encaminhamentos

Este produto descreve a tentativa de iniciar a implementação da interconexão dos repositórios digitais da Funarte, em busca de prosseguir com a agregação dos acervos propostas no primeiro produto do projeto. Contudo, foram encontradas barreiras principalmente quanto aos acervos alocados nos repositórios de Biblioteca e Acervo do sistema SophiA. Abaixo são listados os resultados do diagnóstico de implementação da interconexão de repositórios para cada um dos acervos geridos pelo CEDOC.

- **Catálogo CEDOC (SophiA Biblioteca):** O acesso ao OAI-PMH via URL foi efetivado, porém não foram encontrados padrões de metadados mapeados e, devido ao fato de que para a coleta de registros ser obrigatório a indicação de um padrão de metadados, não foi possível realizar a coleta dos dados.
- **Acervos Privados (AtoM):** Não foram encontrados problemas na forma de conexão com o sistema através do protocolo OAI-PMH, contudo, até a data de execução dessa etapa do projeto o acervo ainda estava incompleto.
- **Acervo Sérgio Britto Digital (SophiA Acervo):** Não foram identificadas possibilidades de conexão com o sistema através do OAI-PMH. Como alternativa encaminhada, optou-se pela realização da migração do acervo para o Tainacan, que já foi efetivada, o que permite então a criação da página de exibição do acervo para compor a área do CEDOC Digital.

Diante da impossibilidade de manter a proposta original de interconectar os repositórios através do protocolo OAI-PMH, surge a alternativa de compor um ambiente agregado de uma parcela do acervo do CEDOC através do Tainacan, o CEDOC Digital.

O CEDOC Digital foi proposto pela equipe do CEDOC, e prevê a criação de uma área no site da Funarte, dentro do contexto de páginas do CEDOC, que exibam o Acervo Sergio Britto Digital, já migrado para o Tainacan, e uma parcela do acervo de Arquivos Privados, que já existiam no Tainacan devido a um processo de migração anterior.

Assim, ainda continuariam como estão o *Catálogo CEDOC* e o acervo de Arquivos Privados, disponíveis para consulta de acesso do usuário através da interface web de seus sistemas, e o CEDOC Digital será uma área em que tanto o acervo Sérgio Brito Digital quanto o conjunto de objetos do acervo de Arquivos Privados serão apresentados em exposições on-line através das funcionalidades de criação de páginas com objetos do acervo que o Tainacan possui.

Dessa forma, tem-se como encaminhamentos para os próximos produtos do projeto:

- Estruturar uma proposta sobre a construção da área do CEDOC Digital.
- Formação em Tainacan para a equipe do CEDOC.



## Referências

FUNARTE, Centro de Documentação da Funarte – Biblioteca Edmundo Moniz. 2021. Disponível em: <<https://www.funarte.gov.br/espaco-cultural/centro-de-documentacao-da-funarte-biblioteca-edmundo-moniz/>>. Acesso em: 25 de fev. de 2021.

PORTAL SOPHIA, Sobre o SophiA Biblioteca. 2021. Disponível em: <<http://www.portalsophia.com.br/SobreBiblioteca.aspx>>. Acesso em: 24 de fev. de 2021

IBICT, Potocolos OAI-PMH / OAI-ORE. 2021. Disponível em: <[http://wiki.ibict.br/index.php/Potocolos\\_OAI-PMH\\_/OAI-ORE](http://wiki.ibict.br/index.php/Potocolos_OAI-PMH_/OAI-ORE)>. Acesso em: 25 de fev. de 2021.