



Relatório Técnico 1

Diagnóstico do acervo da Funarte

Projeto de Pesquisa: Pesquisa e desenvolvimento para organização da informação aos novos modelos de reuso de conteúdos digitais



Março 2019

Apresentação

Este Projeto de Pesquisa é uma iniciativa conjunta do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e da Fundação Nacional de Artes (Funarte). Seu objetivo principal é a produção e a implementação de uma nova estratégia digital para os acervos e informações de relevância da Funarte, visando a reformulação de seu portal de informações e repositório digital.

Os resultados esperados para este projeto são:

1. Diagnóstico do acervo da Funarte;
2. Repositório digital da Funarte migrado por meio de estratégia de criação e metodologia, com foco nos acervos já digitalizados no CEDOC;
3. Portal Funarte Digital;
4. Produções teóricas e demais atividades de pesquisa, como *workshops*, voltados para a área da Ciência da Informação.

O presente relatório apresenta todas as atividades executadas para se atingir o **Resultado 1: Diagnóstico do acervo da Funarte.**



Sumário

1. Introdução.....	4
2. Resultados Descritivos.....	5
2.1 Arquivos privados.....	5
2.2 Audiovisual.....	7
2.3 Biblioteca.....	8
2.4 Site.....	9
3. Considerações.....	11
ANEXOS.....	12
Anexo I - Metodologia de extração dos dados dos catálogos (Word e PDFs).....	12
Anexo II - Apresentação síntese do diagnóstico do acervo CEDOC.....	17



1. Introdução

O diagnóstico do acervo Funarte foi a primeira ação do projeto envolvendo o processo de migração do acervo do CEDOC para a plataforma Tainacan. Para isso uma visita foi realizada entre os dias 12 e 13 de Julho de 2018 no CEDOC como objetivo de entender as práticas de preservação, e conhecer a dinâmica de trabalho. Possibilitando assim, analisar as potenciais necessidade informacionais do departamento, com conversas feitas diretamente com as pessoas que organizam, produzem e gerenciam as informações e seus sistemas.

Ao longo dos dois dias de visita, foram feitas reuniões com as coordenações e com parte dos gestores da documentação do CEDOC, em rodas de conversa que tinham como objetivo entender as práticas informacionais vigentes, e a lógica de funcionamento institucional que contextualiza a manutenção dos documentos.

Em parceria com os responsáveis de cada área foram visitados os acervos físicos, de cartazes, publicações da Funarte e arranjos arquivísticos sobre artistas e peças teatrais, além disso, acessados os acervos armazenados tanto no sistema interno quanto aqueles disponíveis via SophiaWeb, e também a área de gestão dos arquivos permanentes.

De acordo com a quantidade de material apresentado, um grande número de pastas com arquivos de catálogo de peças teatrais, arquivos de texto, PDFs e planilhas com a relação de itens dos arranjos, foi realizada uma coleta guiada para posterior análise do material e identificação aproximada do tamanho do acervo completo do CEDOC.

Além dos arquivos distribuídos no sistema de pastas interno, junto com a equipe de suporte tecnológico, que gerencia os bancos de dados do CEDOC, foi realizada a coleta das duas bases de dados dos acervos on-line Sergio Britto e do SophiaWeb com conteúdo da biblioteca. Como o formato das bases de dados disponibilizado foi o SQL Server, foi necessário, ainda no momento da visita efetuar a transformação para o formato MySQL, que permitiu a confirmação dos dados obtidos, e a validação conjunta com a equipe que gerencia a base de dados.

De posse dos dados obtidos nos diversos formatos na visita realizada no CEDOC, a equipe de Ciência de Dados do Laboratório de Políticas Públicas Participativas efetuou um processo de varredura de catálogos, inventários parciais e dos quantitativos das bases de dados coletadas, em busca de fornecer uma análise descritiva aproximada das características quantitativas de todo o acervo disponível no CEDOC.

2. Resultados Descritivos

Para melhor descrever os resultados obtidos dos esforços de mapeamento quantitativo do acervo do CEDOC através dos arquivos coletados no primeiro encontro presencial, quatro dimensões podem ser expressar para apontar a estrutura dos arquivos, o arquivo privado, o audiovisual, a biblioteca e o site.

Os seguintes totais aproximados foram identificados na primeira visita:

- Arquivo Privado (116.379 – 30,66% Itens):
 - 3.467 pastas – 84.684 Documentos – 28.228 Folhas
- Audiovisual (117.404 – 30,93% Itens):
 - 14.369 Cartazes – 98.060 Fotos – 810 Mídias – 4165 Referências
- Biblioteca (135.011 – 35,57% Itens):
 - 191 Fotos – 220 Documentos – 134.600 Registros (SophiaWeb Biblioteca)
- Site (10.753 – 2,83% Itens):
 - 22 Livros – 4.861 Documentos – 343 Programas – 1.200 Partituras – 70 Arquivos Musicais 52 Projetos – 4.205 Registros (SophiaWeb Sérgio Britto)

Resultado: um total aproximado de 379.547 itens, e uma proporção de 38.35% de itens digitalizados no site e no SophiaWeb.

2.1 Arquivos privados

Ao analisar o conjunto dos arquivos privados, 45 foram identificadas separadamente:

Acervo José Renato
Adamastor_CAMARÁ
ARQUIVO HENRIQUE PONGETTI
ARQUIVO IVAN SENNA
ARQUIVO JACKSON DE SOUZA
ARQUIVO JORACY CAMARGO
ARQUIVO JUSSARA PORTILHO LINS
ARQUIVO LENA BRASIL
ARQUIVO LUIS CARLOS RIPPER
ARQUIVO MANOEL BERLINCK
ARQUIVO MARCONDES MESQUEU
ARQUIVO MARTINEZ CORREA
ARQUIVO NILSON CONDÉ
ARQUIVO NORMA GERALDY
ARQUIVO OSCAR FELIPE
ARQUIVO PAOLETTI
Arquivo Reginaldo Saddi
ARQUIVO YARA AMARAL
Aurimar_ROCHA
COLEÇÃO ANA AUGUSTA DE ALMEIDA
COLEÇÃO ARACY CORTES
COLEÇÃO BEMVINDO SIQUEIRA
Coleção Carlos José Villar

COLEÇÃO DÉCIO L. P. ALVES
Coleção Edgard da Rocha Miranda
COLEÇÃO EDIGAR DE ALENCAR
Coleção Estellita Bell
Coleção Fabio Sabag
COLEÇÃO GRANDE OTELO
COLEÇÃO IVETTE AMARAL
COLEÇÃO JORGE GONÇALVES
Coleção Klaus Vianna
COLEÇÃO LUIS DE LIMA
COLEÇÃO MARIA HELENA ARANHA
COLEÇÃO MARIA IZABEL MINVIELLE
Coleção Maria Wanderley
COLEÇÃO MARIO AIZEN
COLEÇÃO MOEMA TOSCANO
COLEÇÃO ORLANDO CODÁ
COLEÇÃO PAULO ROBERTO FERREIRA
Coleção Rodolfo Mayer
Lista de pastas - Aracy Cortes
Lista de pastas - Eva Todor
Lista de pastas - Luis de Lima
Lista de pastas - Roberto Pontual

Essas coleções foram identificadas a partir do nome das pastas nos arquivos coletados, onde cada uma das pastas tem o nome de um artista, e dentro das pastas conjuntos de arquivos que expressam catálogos e itens sobre aquele artista de posse do CEDOC, formando um arranjo de coleções que estrutura o contexto dos arquivos privados.

Ao fazer a verificação e análise das 45 coleções 214 itens ao total foram identificados e categorizados, sendo tipo de item em maior quantidade de folhas avulsas (63.41%), seguido de recortes de jornais (4,25%) entre outros itens:

Tipo Geral	Nº de Documentos	%	Tipo Geral	Nº de Documentos	%	Tipo Geral	Nº de Documentos	%
Folhas	67751	63,41%	Saúde	113	0,11%	Documentos de Educação	15	0,01%
Recorte de Jornal	4543	4,25%	Recorte de Revista	109	0,10%	Cartão	14	0,01%
Clipping	3701	3,46%	Documentos Fiscais	100	0,09%	Listas	14	0,01%
Documentos Diversos	2907	2,72%	Notas Fiscais	96	0,09%	Poster	14	0,01%
Texto - Peça Teatral	2451	2,29%	Documentos de Teatro	93	0,09%	Trabalho Acadêmico	14	0,01%
Fotografia	2411	2,26%	Discos	91	0,09%	Planta	13	0,01%
Documentos Não Discriminado	2406	2,25%	Panelto	91	0,09%	Solicitação	13	0,01%
Recortes	2031	1,90%	Relações	91	0,09%	Contabilidade	12	0,01%
Partitura	2008	1,88%	Cartões	90	0,08%	Telegrama	12	0,01%
Programa	1931	1,81%	Material de divulgação	85	0,08%	Fita K7	11	0,01%
Dossiê	1485	1,39%	Rascunhos manuscritos	84	0,08%	Rascunhos	11	0,01%
Correspondência	1341	1,26%	Recibo	79	0,07%	Catálogo	10	0,01%
Livros	873	0,82%	Fragmento de Roleiro	77	0,07%	Documento de Carnaval	10	0,01%
Argumento	788	0,74%	Recorte de revista e periódicos	73	0,07%	Fitas Magnéticas	10	0,01%
Poemas e letras de música	610	0,57%	Fragments - Peça Teatral	63	0,06%	Jornal	10	0,01%
Periódico	494	0,46%	Croquis	59	0,06%	Programa de Espetáculo	10	0,01%
Recortes de Periódico	484	0,45%	Esporle	54	0,05%	Carteira de Identidade	9	0,01%
Caixa	442	0,41%	Roupas	50	0,05%	Comunicado	9	0,01%
Textos	427	0,40%	Textos diversos	49	0,05%	Entidades Teatrais	9	0,01%
Pastas	404	0,38%	Homenagem	43	0,04%	Folha de Acetato	9	0,01%
Roteiro	346	0,32%	Disco de Vinil	42	0,04%	Fragments	9	0,01%
Programas/Panfletos	337	0,32%	Ingressos	41	0,04%	Slides	9	0,01%
Variiedades	301	0,28%	Albums de Fotografia	40	0,04%	Talão de Cheques	9	0,01%
Comprovação profissional	299	0,28%	Placas	40	0,04%	Envelopes	8	0,01%
Grações pessoais	297	0,28%	Arte	38	0,04%	Lista	8	0,01%
Documentos de Terceiros	284	0,27%	Contos	37	0,03%	Peças	8	0,01%
Programa - Peça Teatral	273	0,26%	Jornais	37	0,03%	Requisições	8	0,01%
Folhetos	266	0,25%	Contrato	35	0,03%	Cadastro	7	0,01%
Documentos Pessoais	256	0,24%	Memorando	34	0,03%	Catas	7	0,01%
Letras de Músicas	247	0,23%	Documentos de Produção	30	0,03%	Ficha Técnica	7	0,01%
Cartaz	220	0,21%	Documentos sobre Filmes	28	0,03%	Produção Peça Teatral	7	0,01%
Fragmento de Texto	169	0,16%	Curriculo	25	0,02%	Rascunhos Diversos	7	0,01%
Patrimônio	148	0,14%	Dossiê/Prêmio	24	0,02%	Texto	7	0,01%
Anotações	144	0,13%	Registros biográficos	23	0,02%	Album	6	0,01%
Peças Teatrais	144	0,13%	Boletim	22	0,02%	Balancete	6	0,01%
Poesias	140	0,13%	Calendário	21	0,02%	Cumpom	6	0,01%
Revista	139	0,13%	Carteira	21	0,02%	Editais	6	0,01%
Produções de Terceiros	138	0,13%	Poemas	21	0,02%	Leis	6	0,01%
Cartas	137	0,13%	Desenho	20	0,02%	Prêmio	6	0,01%
Notas	131	0,12%	Nota fiscal	19	0,02%	Apostila	5	0,00%
Convite	120	0,11%	Textos de Romances	19	0,02%	Certidões	5	0,00%
			Livro	18	0,02%	Diploma	5	0,00%
			Bilhetes	17	0,02%	Escritos para Jornais	5	0,00%

A situação dos itens presentes na dimensão de Arquivos Privados indica catálogos sobre diversos artistas, dando uma noção aproximada dos itens presentes no acervo do CEDOC, porém nem todos estes itens identificados estão digitalizados, e existe certa dificuldade de identificar os itens presentes também no Sophia Web ou no Site devido a falta de números identificadores dos itens.

2.2 Audiovisual

Os arquivos da dimensão do audiovisual, diferentemente da dimensão dos arquivos privados, não se encontram separados por pastas em arranjos de coleções, e sim são um conjunto de arquivos textuais, PDFs e planilhas com catálogos parciais da quantidade de documentos por tipo de suporte, sendo a maioria dos documentos fotos (86,22%):

Tipo Geral	Nº de Documentos	%
Fotos	98060	86,22%
Cartazes	14369	12,63%
Mídias Digitalizadas	810	0,71%
Referências	487	0,43%

Além disso, alguns catálogos apontam a informação do tipo de mídia em que determinado documento ou arquivo se encontra revela que a maioria deles está no formato de Fita K7 (41,60%):

Tipo de Mídia	Nº de Documentos	%
Fitas K7	337	41,60%
UMATICS	215	26,54%
Betacam	173	21,36%
VHS	85	10,49%

As condições dos itens do acervo audiovisual seguem a mesma condição dos itens dos arquivos privados, o que existe de documentação são catálogos dos documentos físicos existentes do CEDOC, e não existe uma certeza sobre quais desses documentos estão digitalizados ou não, dificultando o processo de migração dos acervos para o Tainacan.

2.3 Biblioteca

Na dimensão dos itens de Biblioteca, foram verificados 3 itens, o catálogo de publicações da Funarte com capa, os livros digitalizados, e os registros do SophiaWeb. As quantidades de itens existentes no Sophia Web foram retiradas dos relatórios do sistema e apontam como modo de aquisição mais representativos dos itens obtidos pelo CEDOC a Doação (60,86%):

Registros SophiaWeb

Aquisição	Nº de Documentos	%
Doação	81919	60,86%
Outros	29547	21,95%
Compra	17223	12,80%
Não Especificado	5895	4,38%
Permuta	16	0,01%
Total	134600	100,00%

2.4 Site

Ao verificar os itens disponíveis no site, a maioria deles correspondem a documentos (45,21%) e registros (39,21%):

Tipo Geral	Nº de Documentos	%
Documentos	4861	45,21%
Registros	4205	39,11%
Partituras	1200	11,16%
Programas	343	3,19%
Arquivos Musicais	66	0,61%
Projetos	52	0,48%
Livros On-line	22	0,20%
Áudios	4	0,04%
Total	10753	100,00%

Além dos itens distribuídos pelo site, o portal Brasil Memória das Artes apresentou um total de 9.674 itens, distribuídos em 14 coleções, representando 2,56% do total de itens verificados até então:



Coleção	Link	Vídeo	Fotografia	Textos	Áudio	Total	%
Projeto Pinguinha	http://www.funarte.gov.br/brasilmemoriadasartes/acervos/#pinguinha	88	301	212	783	1384	14,31%
Sala Funarte	http://www.funarte.gov.br/brasilmemoriadasartes/acervos/#serie-cartazes-da-sala-funarte	14	154	67	869	1104	11,41%
Discos PRO-MEMUS	http://www.funarte.gov.br/brasilmemoriadasartes/acervos/#discos-pro-memus	0	165	28	337	530	5,48%
Discos Projeto Almirante	http://www.funarte.gov.br/brasilmemoriadasartes/acervos/#discos-projeto-almirante	21	189	41	216	467	4,83%
Atores do Brasil	http://www.funarte.gov.br/brasilmemoriadasartes/acervos/#atores-do-brasil	14	366	38	26	444	4,59%
Cenário e figurino	http://www.funarte.gov.br/brasilmemoriadasartes/acervos/#cenario-e-figurino	7	193	18	8	226	2,34%
Foto Carlos	http://www.funarte.gov.br/brasilmemoriadasartes/acervos/#foto-carlos	14	126	24	3	167	1,73%
Nelson Rodrigues	http://www.funarte.gov.br/brasilmemoriadasartes/acervos/#nelson-rodrigues	28	72	33	0	133	1,37%
INFoto	http://www.funarte.gov.br/brasilmemoriadasartes/acervos/#irifoto	8	40	73	0	121	1,25%
Série Depoimentos	http://www.funarte.gov.br/brasilmemoriadasartes/acervos/#serie-depoimentos	7	47	12	16	82	0,85%
Paschoal Carlos Magno	http://www.funarte.gov.br/brasilmemoriadasartes/acervos/#paschoal-carlos-magno	0	73	5	0	78	0,81%
Família Vianna	http://www.funarte.gov.br/brasilmemoriadasartes/acervos/#familia-vianna	0	31	11	0	42	0,43%
Augusto Boal	http://www.funarte.gov.br/brasilmemoriadasartes/acervos/#augusto-boal	7	20	11	1	39	0,40%
Labanca	http://www.funarte.gov.br/brasilmemoriadasartes/acervos/#labanca	0	18	2	0	20	0,21%
Walter Pinto	http://www.funarte.gov.br/brasilmemoriadasartes/acervos/#walter-pinto	0	0	0	0	0	0,00%
Total						9674	100,00%

3. Considerações

A partir da visita presencial, onde ocorreram reuniões de apresentação da equipe, das formas de curadoria do acervo do CEDOC e a coleta do material para dimensionamento, foi possível entender a estrutura de arquivos obtida, bem como a estrutura do site e das bases de dados coletadas, etapa fundamental do diagnóstico do acervo, o que permitiu a produção de análises descritivas que apontassem aproximadamente quantidade total de itens, e a quantidade digitalizada que tem potencial de migração para o Tainacan.

Foi identificado assim o total aproximado de 379.547 mil itens, divididos em quatro dimensões estruturais propostas, o Acervo Privado, o Audiovisual, a Biblioteca e o Site, em que somente esses últimos dois apresentam capacidade de migração por já estarem digitalizados, os demais apontam a organização catalográfica dos itens disponíveis em acervo físico, o que permitiu indicar o potencial de digitalização e difusão do acervo do CEDOC em futuras estratégias de digitalização.

Para efetuar esse trabalho de mensuração, o laboratório desenvolveu metodologias de extração e conversão de dados (Anexo I), reutilizando as informações contidas em PDFs ou documentos de texto em planilhas, o que permitiu a melhor manipulação e organização dos dados para a mensuração, e futuras estratégias de conversão dos dados para complementar os metadados dos itens digitalizados na migração para o Tainacan.

Por fim, esses dados foram reunidos em uma apresentação, e utilizados em reuniões com a equipe e coordenação do CEDOC, para traçar quais os acervos prioritários para a migração inicial, e quais os projetos de digitalização vigentes e já concluídos que otimizem o processo de criação de coleções e itens no Tainacan.

ANEXOS

Anexo I - Metodologia de extração dos dados dos catálogos (Word e PDFs)

Foi desenvolvido e adotado um conjunto de práticas para lidar com documentos com essas características, a partir da observação do trabalho desenvolvido no laboratório, que envolve quatro etapas:

1. Etapa PDF/DOCS para texto editável:

- ✓ Análise da estrutura do conteúdo

O processo de extração de conteúdo de documentos em Word ou PDF se inicia na avaliação do conteúdo presente no documento e sua eventual sistematização. Isto significa procurar padrões comuns em relação ao preenchimento e separação das informações o que, em uma planilha, significará em colunas que irão receber o conteúdo do documento.

- ✓ Análise do conteúdo

O passo seguinte consiste na análise do conteúdo e a formatação de uma proposta de colunas da planilha que receberá a conversão do documento. Extração bruta via Google Docs ou Word

No caso de arquivos do tipo PDF, a abertura e conversão para o formato de texto editável pode ser realizada pela importação pelo Word ou pelo Google Docs:

No word, basta clicar no arquivo .pdf abrindo seu navegador de contexto (pelo windows ou mac) e selecionar “Abrir com > Word” e o arquivo será convertido para texto.

Já para conversão com o Google Docs, é necessário enviar o arquivo PDF para o Google Drive e, pelo navegador, abrir a visualização do arquivo e, em seguida, selecionar no topo da janela “Abrir com > Documentos Google”. A conversão será realizada em uma nova aba e em alguns instantes estará concluída.

A principal diferença entre o uso do Word e do Google Docs está na interpretação das quebras de linha feita entre os programas. Geralmente, o Google Docs tende a ser capaz de identificar melhor conjunto de textos, o que implica em um resultado mais favorável para a transposição do conjunto textual para outros programas.

- ✓ **Sistematização dos metadados:** Identificado padrões de colunas e intervalos

A análise da estrutura e do conteúdo é um processo que pode e deve ser realizado algumas vezes consecutivas, pois consiste no planejamento e na base do trabalho de extração a ser desenvolvido. Sua má formatação e/ou a ausência de colunas previstas na base em questão pode gerar a necessidade de retrabalho.

Uma vez determinada a criação das colunas que irão compor a coluna, é necessário identificar no conjunto do documento os eventuais separadores, tais como vírgulas, dois pontos, ponto e vírgula, aspas ou termos que antecedem de forma padronizada o conteúdo como “Título”, “Nome”, “Número de registro” e afins. Ou seja, a identificação de elementos que possam servir para a sistematização de um conjunto de valores para a tabulação do conjunto textual.

2. Base de dados “bruta”

Com base nas interpretações sobre o material nos passos anteriores, já é possível quebrar o conjunto de dados, isto pode ser realizado a partir da separação de caracteres pelo Excel ou também pelo Open Refine.

A documentação específica ou tutoriais para cada software podem ser consultadas nestes links:

- **Limpeza de dados con OpenRefine**

<https://programminghistorian.org/es/lecciones/limpieza-de-datos-con-OpenRefine>

- **Limpeza de Dados com o Refine**

<http://geojournalism.org/pt/2013/07/cleaning-data-with-refine/>

- **Para expandir o conteúdo da célula para múltiplas células**

https://help.libreoffice.org/Calc/Text_to_Columns/pt-BR

- **Separar Campos tendo como referência espaço em branco**

<https://ask.libreoffice.org/pt-br/question/44968/separar-campos-tendo-como-referencia-espaco-em-branco/>

- **Dividir o texto em diferentes colunas com o Assistente para conversão de texto em colunas**

<https://support.office.com/pt-br/article/dividir-o-texto-em-diferentes-colunas-com-o-assistente-para-convers%C3%A3o-de-texto-em-colunas-30b14928-5550-41f5-97ca-7a3e9c363ed7>

3. Etapa EXCEL

- Analisar os dados transpostos:

Observando a ausência de conteúdo, a integridade do volume de dados (pode ser realizado conferindo a quantidade de caracteres entre o documento original e a transposição).

- Preservar o original

Como boa prática, recomenda-se preservar a transposição criando uma cópia para realizar as próximas etapas. Isso garante um “backup” caso futuras ações não saiam como esperado e também como um local de acesso rápido para conferência e comparação dos dados tratados.

- Começar extraindo os dados pelas pontas. De forma que os dados centrais fiquem por último.
- Funções:
 - Procurar
 - Esquerda
 - Direita
 - Macro remove character from string
 - First
 - Last
 - Ext.Texto
- Aplicar filtro
- Verificar distorções
- Replicar os passos por lote até completar

4. Etapa Open Refine

Uma vez terminado o trabalho de transposição dos dados para a planilha é necessário o trabalho de revisão do conteúdo das colunas, tanto para garantir que os campos estejam corretamente preenchidos, como também para normalização das informações.

Este passo a passo pode ser realizado a partir da visualização por filtros na planilha, o que permite a comparação de valores caso a caso. Ou também, pode ser

utilizado o OpenRefine como ferramenta para análise e reconciliação de valores em lotes.

A instalação e documentação do software está disponível [neste link](#).

- **Importar os dados**

Importe a planilha para o OpenRefine. Como boa prática é recomendado nomear o projeto adequadamente.

Garanta que a importação preserve a estrutura de dados e tabulação original, bem como selecione o formato de caracteres adequado ao idioma (geralmente, UTF-8)

- **Efetuar limpeza de espaços iniciais, espaços consecutivos e finais das células**

Com o projeto aberto, realize a limpeza de espaços iniciais e finais com o comando: Clique na “seta para baixo” à esquerda do nome de cada coluna > No menu, selecione “Editar Células” > Transformações Comuns > Em seguida clique em “Limpar espaços no início e no final dos valores”.

Para a limpeza de espaços consecutivos:

Clique na “seta para baixo” à esquerda do nome de cada coluna > No menu, selecione “Editar Células” > Transformações Comuns > Em seguida clique em “Limpar espaços consecutivos”.

Repita a operação em cada coluna.

Este processo garante que a base contenha menos divergências entre os valores de cada coluna.

Neste mesmo menu, também é possível determinar o tipo de conteúdo da coluna entre “Número”, “Data” ou texto. Como também a manipulação dos conteúdos entre Maiúsculos, minúsculos e maiúsculos no primeiro caractere.

- **Realizar edição de colunas e clusterização**

Para a edição dos termos em massa, execute o seguinte comando:

Clique na “seta para baixo” à esquerda do nome de cada coluna > No menu, selecione “Facetas” > “Faceta de Texto”: Uma janela será exibida na lateral esquerda da janela de trabalho, com isto é possível a edição de termos em lote.

- **Identificar similares e variações**

Para a normalização de valores por comparação, acesse:

Clique na “seta para baixo” à esquerda do nome de cada coluna > No menu, selecione “Clusterizar e Editar” .

Será exibida uma janela acima da planilha de trabalho para a seleção de termos similares. Clique na caixa de seleção para valores equivalentes.

Também é possível editar o conjunto de valores no campo de texto à esquerda da caixa de seleção.

Para mais operações no menu de clusterização, acesse a documentação do OpenRefine disponível neste [link](#).

- **Re-clusterizar**

Repita o processo anterior com diferentes metodologias para garantir mais conformidade aos valores da base. Também execute este processo em todas as colunas.

- **Exportar**



Laboratório de Políticas
Públicas Participativas

Anexo II - Apresentação síntese do diagnóstico do acervo CEDOC

FUNARTE

Análise Quantitativa dos Arquivos e
Catálogos - CEDOC
10/09/2018

1

Situação inicial – Primeira Visita

- Visitamos o Centro de Informação e Documentação –CEDOC nos dias 12 e 13 de Julho;
- O objetivo foi explorar a situação documental do departamento, entender as práticas de preservação, e conhecer a dinâmica de trabalho.
- Essa primeira etapa nos permite analisar as potenciais necessidades informacionais do departamento, com conversas feitas diretamente com as pessoas que organizam, produzem e gerenciam as informações e seus sistemas.

2

Situação inicial – Entrevistas

- Uma das etapas foi reunir as coordenações e parte do pessoal que gerencia a documentação do CEDOC, fazer uma roda de conversa e entender quais as práticas informacionais e a lógica de funcionamento da instituição.
- Foram visitados os acervos físicos e acessada a parte dos acervos digitalizada no SophiaWeb, e do sistema de gestão dos arquivos permanentes.

3

Situação inicial – Coleta

- Ao observar uma grande quantidade de arquivos com característica de inventário de pastas e documentos dos artistas e peças teatrais, com a ajuda das coordenações, foi possível coletar várias pastas com PDFs, arquivos textuais, e planilhas com informações sobre os itens dos acervos.
- Além disso, também foram recuperados os backups de banco de dados dos acervos digitalizados no SophiaWeb, da Biblioteca, e do Acervo Sérgio Britto.

4

Análise Quantitativa – Arquivos

- De posse dos arquivos coletados na primeira visita ao CEDOC, bem como do Backup dos Banco de Dados, já transformados para o formato MySQL por nós no segundo dia, a equipe de ciência de dados do laboratório analisou todas as pastas e arquivos, além dos registros das bases de dados e os acervos disponíveis no site, de forma quantitativa, reunindo os dados em uma noção aproximada da quantidade de itens digitalizados e físicos existentes.
- Assim, é possível entender a escala e o estado da massa informacional que se tem. Permitindo tomar decisões estratégicas sobre os acervos,

5

Resultados

6

Resultados - Dimensões

- **Foram divididas 3 grandes dimensões de acordo com os arquivos:**
 - **Arquivo Privado (116.379 – 30,66% Itens):**
 - 3.467 pastas – 84.684 Documentos – 28.228 Folhas
 - **Audiovisual (117.404 – 30,93% Itens):**
 - 14.369 Cartazes – 98.060 Fotos – 810 Mídias – 4165 Referências
 - **Biblioteca (135.011 – 35,57% Itens):**
 - 191 Fotos – 220 Documentos – 134.600 Registros (SophiaWeb Biblioteca)
 - **Site (10.753 – 2,83% Itens):**
 - 22 Livros – 4.861 Documentos – 343 Programas – 1.200 Partituras – 70 Arquivos Musicais – 52 Projetos – 4.205 Registros (SophiaWeb Sérgio Britto)

Total Aproximado de Itens: 379.547

Percentual Aproximado Digitalizado (Site e SophiaWeb) – 38,35%

7

Resultados - Arquivo Privado

- Foram verificadas 45 coleções:

Acervo José Renato	COLEÇÃO DÉCIO L. P. ALVES
Ademastor_CAMARÁ	Coleção Edgard da Rocha Miranda
ARQUIVO HENRIQUE PONGETTI	COLEÇÃO EDIGAR DE ALENCAR
ARQUIVO IVAN SENNA	Coleção Estelita Bell
ARQUIVO JACKSON DE SOUZA	Coleção Fabio Sabaq
ARQUIVO JORACY CAMARGO	COLEÇÃO GRANDE OTELO
ARQUIVO JUSARA PORTILHO LINS	COLEÇÃO IVETTE AMARAL
ARQUIVO LENA BRASIL	COLEÇÃO JORGE GONÇALVES
ARQUIVO LUIS CARLOS RIPPER	Coleção Klaus Vianna
ARQUIVO MANOEL BERLINCK	COLEÇÃO LUIS DE LIMA
ARQUIVO MARCONDES MESQUITA	COLEÇÃO MARIA HELENA ARANHA
ARQUIVO MARTINEZ CORREA	COLEÇÃO MARIA IZABEL MINVIELLE
ARQUIVO NILSON CONDÉ	Coleção Maria Wanderley
ARQUIVO NORMA GERALDY	COLEÇÃO MARIO AIZEN
ARQUIVO OSCAR FELIPE	COLEÇÃO MOEMA TOSCANO
ARQUIVO PAOLETTI	COLEÇÃO ORLANDO CODÁ
Arquivo Reginaldo Saddy	COLEÇÃO PAULO ROBERTO FERREIRA
ARQUIVO YARA AMARAL	Coleção Rodolfo Mayer
Aurimar_ROCHA	Lista de pastas - Aracy Cortes
COLEÇÃO ANA AUGUSTA DE ALMEIDA	Lista de pastas - Eva Todor
COLEÇÃO ARACY CORTEZ	Lista de pastas - Luis de Lima
COLEÇÃO BEMVINDO SIQUEIRA	Lista de pastas - Roberto Pontual
Coleção Carlos José Villar	

8

Resultados - Arquivo Privado

- Das 45 coleções foram classificados 214 tipos diferentes de documentos:

Tipo Geral	Nº de Documentos	%	Tipo Geral	Nº de Documentos	%	Tipo Geral	Nº de Documentos	%
Folhas	87781	83,41%	Símbolo	110	0,11%	Documentos de Educação	13	0,01%
Recorte de Jornal	4543	4,29%	Recorte de Revista	109	0,10%	Cartão	14	0,01%
Cipping	3701	3,48%	Documentos Fiscais	100	0,09%	Lista	14	0,01%
Documentos Diversos	2907	2,72%	Notas Fiscais	96	0,09%	Postal	14	0,01%
Texto - Peça Teatral	2451	2,29%	Documentos de Teatro	93	0,09%	Trabalho Acadêmico	14	0,01%
Fotografia	2411	2,28%	Discos	91	0,09%	Plano	13	0,01%
Documentos Não Discriminados	2406	2,29%	Partido	91	0,09%	Solicitação	13	0,01%
Recortes	2031	1,90%	Relações	91	0,09%	Constituição	13	0,01%
Partitura	2008	1,89%	Cartões	90	0,08%	Telegrama	12	0,01%
Programa	1931	1,81%	Material de divulgação	85	0,08%	Foto RT	11	0,01%
Crash	1465	1,39%	Rascunhos manuscritos	84	0,08%	Rascunhos	11	0,01%
Correspondência	1341	1,26%	Rádio	79	0,07%	Catálogo	10	0,01%
Livros	873	0,82%	Folheto de Folclore	77	0,07%	Documento de Carnaval	10	0,01%
Argumento	785	0,74%	Recorte de análise e pedidos	73	0,07%	Fitas Magnéticas	10	0,01%
Poemas e letras de músicas	610	0,57%	Folhetos - Peça Teatral	63	0,06%	Jornal	10	0,01%
Periódico	494	0,46%	Croquis	59	0,06%	Programa de Espetáculo	10	0,01%
Recortes de Periódico	484	0,45%	Esposo	54	0,05%	Carteira de Identidade	9	0,01%
Cartão	442	0,41%	Roupas	50	0,05%	Comunicado	9	0,01%
Teatro	427	0,40%	Textos diversos	49	0,05%	Entidades Teatrais	9	0,01%
Passes	404	0,38%	Homenagens	43	0,04%	Folha de Acató	9	0,01%
Rolário	348	0,32%	Disco de Vinil	42	0,04%	Folhetos	9	0,01%
Programas/Periódicos	337	0,32%	Ingressos	41	0,04%	Símbolo	9	0,01%
Variedades	301	0,28%	Álbuns de Fotografia	40	0,04%	Talão de Cheques	9	0,01%
Comprovação profissional	269	0,25%	Placas	40	0,04%	Envelopes	8	0,01%
Concepções pessoais	257	0,24%	Arte	38	0,04%	Lista	8	0,01%
Documentos de Teatros	254	0,24%	Contos	37	0,03%	Peças	8	0,01%
Programa - Peça Teatral	273	0,26%	Jornais	37	0,03%	Requisições	8	0,01%
Folhetos	288	0,27%	Contos	35	0,03%	Cadastro	7	0,01%
Documentos Presenciais	296	0,28%	Manoando	34	0,03%	Cartas	7	0,01%
Letras de Músicas	247	0,23%	Documentos de Produção	30	0,03%	Ficha Técnica	7	0,01%
Cartão	220	0,21%	Documentos sobre Filmes	28	0,03%	Produção Peça Teatral	7	0,01%
Folheto de Teatro	160	0,15%	Curículo	25	0,02%	Rascunhos Diversos	7	0,01%
Platômetro	148	0,14%	Cartão	24	0,02%	Texto	7	0,01%
Anúncios	144	0,13%	Jornais	23	0,02%	Álbuns	6	0,01%
Peças Teatrais	144	0,13%	Registros biográficos	23	0,02%	Balancete	6	0,01%
Poesias	140	0,13%	Boletim	22	0,02%	Compos	6	0,01%
Revista	139	0,13%	Calendário	21	0,02%	Edição	6	0,01%
Produções de Teatros	139	0,13%	Carteira	21	0,02%	Lista	6	0,01%
Cartão	137	0,12%	Poesias	21	0,02%	Prêmio	6	0,01%
Cartão	131	0,12%	Desenho	20	0,02%	Apostila	5	0,00%
Cartão	129	0,11%	Nota Fiscal	19	0,02%	Cartões	5	0,00%
			Textos de Romanos	18	0,02%	Diploma	5	0,00%
			Libro	18	0,02%	Escritos para Jornal	5	0,00%
			Bilhetes	17	0,02%			

Resultados - Audiovisual

- Foram verificados 4 tipos de documentos:

Tipo Geral	N° de Documentos	%
Fotos	98060	86,22%
Cartazes	14369	12,63%
Mídias Digitalizadas	810	0,71%
Referências	487	0,43%

Tipo de Mídia	N° de Documentos	%
Betacam	173	21,36%
VHS	85	10,49%
UMATICS	215	26,54%
Fitas K7	337	41,60%

10

Resultados - Biblioteca

- Na biblioteca foram verificados o catálogo de publicações da Funarte com Capa, os Livros Digitalizados e os registros do SophiaWeb:

Biblioteca	
Tipo Geral	N° de Documentos
Documentos	220
Fotos	191

Registros SophiaWeb		
Aquisição	N° de Documentos	%
Doação	81919	60,86%
Outros	29547	21,95%
Compra	17223	12,80%
Não Especificado	5895	4,38%
Permuta	16	0,01%
Total	134600	100,00%

11

Resultados - Site

- Foram verificados 10.753 documentos provenientes do site:

Tipo Geral	N° de Documentos	%
Documentos	4861	45,21%
Registros	4205	39,11%
Partituras	1200	11,16%
Programas	343	3,19%
Arquivos Musicais	66	0,61%
Projetos	52	0,48%
Livros On-line	22	0,20%
Áudios	4	0,04%
Total	10753	100,00%

12

Resultados- Site – Brasil Memória das Artes

- Foram encontradas 14 coleções no portal Brasil Memória das Artes, a quantidade de documentos ao todo, representa 2,56% do total aproximado de itens encontrados:

Coleção	Link	Video	Fotografia	Textos	Audio	Total	%
Projeto Rodriguez	http://www.funarte.gov.br/brasilmemoriadasartes/acervo/ProjetoRodriguez	88	301	212	782	1384	14,31%
Sala Funarte	http://www.funarte.gov.br/brasilmemoriadasartes/acervo/SalaCartazes-da-Exa-Br-Funarte	14	154	87	869	1104	11,41%
Diário PR CHEMUS	http://www.funarte.gov.br/brasilmemoriadasartes/acervo/Relacoes-pro-memus	0	185	28	337	530	5,48%
Diário Projeto A Imagem	http://www.funarte.gov.br/brasilmemoriadasartes/acervo/Relacoes-pro-ajm	21	189	41	218	467	4,83%
Alcance do Brasil	http://www.funarte.gov.br/brasilmemoriadasartes/acervo/Alcance-do-brasil	14	388	38	28	444	4,58%
Cartão e Ilustração	http://www.funarte.gov.br/brasilmemoriadasartes/acervo/Memoria-e-ilustracao	7	193	18	8	226	2,34%
Foto Cartas	http://www.funarte.gov.br/brasilmemoriadasartes/acervo/Memoria-cartas	14	128	24	3	167	1,73%
Nelson Rodrigues	http://www.funarte.gov.br/brasilmemoriadasartes/acervo/Memoria-rodrigues	28	72	33	0	133	1,37%
Info	http://www.funarte.gov.br/brasilmemoriadasartes/acervo/Info	8	40	73	0	121	1,25%
Série Depoimentos	http://www.funarte.gov.br/brasilmemoriadasartes/acervo/Serie-de-depoimentos	7	47	12	18	82	0,85%
Psicólogo Carlos Magno	http://www.funarte.gov.br/brasilmemoriadasartes/acervo/psicologo-carlos-magno	0	73	5	0	78	0,81%
Família Viana	http://www.funarte.gov.br/brasilmemoriadasartes/acervo/Familia-viana	0	31	11	0	42	0,43%
Augusto Boal	http://www.funarte.gov.br/brasilmemoriadasartes/acervo/Augusto-boal	7	20	11	1	39	0,40%
Labanca	http://www.funarte.gov.br/brasilmemoriadasartes/acervo/Labanca	0	18	2	0	20	0,21%
Vitor Pinó	http://www.funarte.gov.br/brasilmemoriadasartes/acervo/Vitor-pino	0	0	0	0	0	0,00%
Total						9574	100,00%

13

Considerações e Próximos Passos

- **Com os resultados encontrados, temos então uma percepção mais aproximada da massa informacional presente atualmente no CEDOC.**
- **A partir dessa noção, é necessário verificar quais acervos são mais expressivos para a digitalização e disponibilização para acesso on-line.**
- **E também, entender quais estratégias de organização e disseminação podem ser feitas de acordo com a dimensão de cada coleção.**