

**PROJETO INTEROPERABILIDADE ENTRE OS
REPOSITÓRIOS DIGITAIS DO PATRIMÔNIO
CULTURAL BRASILEIRO: DA WEB SEMÂNTICA E
DADOS ABERTOS LIGADOS ÀS FERRAMENTAS DE
BUSCA E RECUPERAÇÃO DA INFORMAÇÃO**

**Relatório referente ao 1º produto do projeto de pesquisa
FAPESP e FCI/UnB**

**Mapeamento sistemático das formas de
organização da informação dos repositórios
digitais de cultura**

**Abril
2020**

1. Introdução	3
1.1. Proposta deste Relatório	3
1.2. Contexto	3
1.3. Instituições Vinculadas à Secretaria Especial da Cultura	5
2. Metodologia	9
2.1. Visão geral (Etapas)	9
3. Variáveis da base de coleta	11
3.1. Variáveis de Identificação	11
3.2. Variáveis de categorização	12
4. Resultados	15
4.1. Números Totais	15
4.2. Pontos de acesso encontrados por variável categórica	17
4.3. Descrição por vinculadas	24
5. Considerações Finais	30
6. Referências	31

1. Introdução

1.1. Proposta deste Relatório

Este documento formaliza o primeiro produto previsto no projeto “Interoperabilidade entre os repositórios digitais do patrimônio cultural brasileiro: da web semântica e dados abertos ligados às ferramentas de busca e recuperação da informação” financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

O primeiro produto, “Mapeamento sistemático das formas de organização da informação dos repositórios digitais de cultura”, almeja compreender os padrões de metadados utilizados por esses repositórios (pontos de acesso encontrados nos portais web das entidades vinculadas à Secretaria Especial da Cultura), as linguagens documentárias para indexação dos itens dos acervos, eventuais regras de catalogação utilizadas, a estrutura organizacional das coleções e a tipologia de acervos documentados.

Com esse resultado espera-se fornecer uma análise das estratégias de organização dos acervos digitais, permitindo a partir deste ponto ser possível pensar e propor formas de agregação dos acervos a por meio de suas sintaxes e semânticas em comum.

1.2. Contexto

Atentos às novas necessidades da sociedade e acompanhando a evolução tecnológica, diversos museus, bibliotecas, arquivos e centros culturais do Brasil e do mundo têm se interessado e buscado caminhos para a digitalização de seus acervos. Para Baggio e Flores (2013), tal movimento também corrobora a necessidade de garantir a preservação digital da informação por meio de uma tecnologia flexível para arquivar, preservar e possibilitar o acesso a documentos em formato digital.

Vale salientar que a digitalização por si só não possibilita o pleno acesso aos acervos. Neste sentido, os repositórios digitais são tendência. Shintaku (2017) explica que, em sua concepção inicial, estes foram vinculados ao “repor”, ou seja, tratavam-se de um local para reposição ou relocação de coisas. Entretanto, o contexto mudou, e atualmente são considerados, com frequência, como primeira fonte.

Um dos pontos fundamentais para a pesquisa científica e o desenvolvimento conceitual e prático das humanidades digitais é a possibilidade de se utilizar bases de dados de ampla representatividade e interesse cultural como fontes de informação para testes de algoritmos, experimentos com ferramentas de análise, aplicação de metodologias quantitativas e qualitativas, comparação com resultados obtidos internacionalmente, além da possibilidade de se testar hipóteses e avançar no conhecimento empírico.

Desse modo, encontrar fontes de informação que possam ser utilizadas como base para pesquisas em humanidades digitais, ou seja, que forneçam informações sobre seus acervos e ainda permitam que essas possam ser reutilizadas, a partir de sua disponibilização em formatos abertos e interoperáveis, torna-se um problema de grande interesse.

Apesar disso, há pouca informação sistematizada no Brasil sobre a existência de fontes de informação na área da cultura e, menos ainda, da qualidade técnica dos dados disponibilizados e se podem ou não serem facilmente reutilizados. Segundo a pesquisa TIC Cultura 2018, do Comitê Gestor da Internet (2019), a disponibilidade de catálogos dos acervos dos arquivos brasileiros na Internet é de 38% das unidades, sendo que 15% dos museus possuem a mesma prática e 12% das bibliotecas. Ou seja, o catálogo dos acervos dessas instituições não está disponível em rede e, muitas das vezes, sequer digitalizados. Tal fato, como se pode avaliar pelos números acima, torna muito difícil o avanço de pesquisas nas ciências humanas, que se aproveitem dos recursos das humanidades digitais. É questão fundamental a ser resolvida para avanço do campo no país.

É, de fato, a perspectiva da digitalização e do reuso desses conteúdos que muitos laboratórios e centros de pesquisa têm se interessado pela possibilidade de utilização dos acervos das instituições memoriais, como ressalta Puntoni (2017, p. 129-130):

“E não podemos deixar de perceber o enorme papel desempenhado neste processo de aproximação da cultura digital e das instituições memoriais pelos centros de pesquisa e pelas universidades, sobretudo pelos profissionais ligados à ciência da informação, aos estudos do patrimônio ou mesmo ao que começava a se definir como humanidades digitais (digital humanities).”

Como forma de problematizar a questão acima e avançar na análise empírica dessas fontes de informação, optou-se nesta pesquisa em focar nas instituições vinculadas à Secretaria Especial da Cultura, anterior Ministério da Cultura, sobretudo por sua relevância e abrangência nacional, sendo elas, de acordo com a Secretaria Especial da Cultura (2013c): o Instituto do Patrimônio Histórico e Artístico Nacional - Iphan, o Instituto Brasileiro de Museus - Ibram, a Agência Nacional do Cinema - Ancine, a Fundação Casa de Rui Barbosa - FCRB, a Fundação Cultural Palmares - FCP, a Fundação Nacional das Artes - Funarte e a Fundação Biblioteca Nacional - FBN.

Por ter como objetivo mapear a organização técnica destes acervos para propor formas de integração dos dados, definiu-se importantes questões a serem respondidas: Quais os tipos das estruturas dos sistemas de informação utilizados nos repositórios digitais? Quais as ferramentas utilizadas para a publicação dos acervos? Quais licenças, padrões de metadados, linguagens documentárias e regras de catalogação são explicitadas na divulgação do acervo? Quais as formas de exposição dos itens dos acervos? Quais as formas de extração dos dados do acervo? Em qual formato os itens do acervo estão disponibilizados? Qual o tamanho do acervo? Quais são os formatos dos arquivos armazenados?

1.3. Instituições Vinculadas à Secretaria Especial da Cultura

O extinto Ministério da Cultura, atual Secretaria Especial da Cultura (2013b), foi criado em 1985 a partir do desmembramento do Ministério da Educação e Cultura e desenvolve ações para reconhecimento da importância da cultura para a construção da identidade nacional.

Desde então, compete a esta secretaria a:

- I. política nacional de cultura;
- II. proteção do patrimônio histórico e cultural;
- III. regulação de direitos autorais;
- IV. assistência e acompanhamento da Casa Civil da Presidência da República e do Instituto Nacional de Colonização e Reforma Agrária - Incra nas ações de regularização fundiária, para garantir a preservação da identidade cultural dos remanescentes das comunidades dos quilombos; e

V. desenvolvimento e implementação de políticas e ações de acessibilidade cultural (SECRETARIA ESPECIAL DA CULTURA, 2013a).

As entidades vinculadas são importantes instituições culturais temáticas que realizam ações e políticas específicas no campo, descritas a seguir.

- **Instituto do Patrimônio Histórico e Artístico Nacional - Iphan**

Em conformidade com Iphan (2014b), o instituto foi criado em janeiro de 1937, por meio da Lei nº 378, para responder pela preservação do Patrimônio Cultural Brasileiro, assegurando sua permanência e usufruto para as gerações presentes e futuras. Além disso, é responsável pela conservação, salvaguarda e monitoramento dos bens culturais brasileiros inscritos na Lista do Patrimônio Mundial e na Lista o Patrimônio Cultural Imaterial da Humanidade.

- **Instituto Brasileiro de Museus - Ibram**

Segundo Ibram (2018b) e Secretaria Especial da Cultura (2013c), o Ibram foi criado em janeiro de 2009, com a assinatura da Lei nº 11.906, para, entre outras finalidades, promover e assegurar a implementação de políticas públicas para o setor museológico, contribuindo para a organização, gestão e desenvolvimento de instituições museológicas e seus acervos.

De acordo com a Secretaria Especial da Cultura (2013c), o Ibram se propõe a estimular a participação de instituições nas políticas públicas para o setor museológico e nas ações de preservação, investigação e gestão do patrimônio cultural musealizado. Conforme Ibram (2018a), o órgão, também responsável pela Política Nacional de Museus, visa o aumento de visitação e arrecadação dos museus, fomento de políticas de aquisição e preservação de acervos e criação de ações integradas entre os museus brasileiros.

O Ibram é responsável pela administração direta de 30 museus, cuja lista completa está disponível no site do Instituto , destes, 14 estão em processo de divulgação do seu acervo na internet, e 16 já tem seu acervo ou parte dele disponibilizado para consulta on-line, são eles: Museu Casa da Princesa, Museu Casa de Benjamin Constant, Museu da Inconfidência, Museu das Bandeiras, Museu das Missões, Museu de Arqueologia de Itaipu, Museu de Arte Sacra da Boa Morte, Museu do Diamante, Museu do Ouro, Museu Histórico Nacional, Museu Regional São João

Del Rey, Museu Regional Casa dos Ottoni, Museu Victor Meirelles, Museu Villa Lobos, Museu Casa da Hera, Museu Casa Histórica de Alcântara.

No ano de 2016, o IBRAM optou por aderir a plataforma Tainacan e vem implementando o sistema em seus museus como estratégia para abertura e disponibilização digital dos acervos musealizados.

- **Agência Nacional do Cinema - Ancine**

Conforme Ancine (2018) e Secretaria Especial da Cultura (2013c), a Ancine foi criada em 2001, pela Medida Provisória 2228-1, e vinculada desde 2003 à atual Secretaria Especial da Cultura. Trata-se de uma agência reguladora cuja missão é desenvolver e regular o setor audiovisual em benefício da sociedade brasileira, além proporcionar o desenvolvimento de uma indústria forte, competitiva e auto-sustentada.

- **Fundação Casa de Rui Barbosa - FCRB**

De acordo com a Fundação Casa de Rui Barbosa (2018b, 2018c), a fundação foi instituída em 1928 e sua missão é desenvolver a cultura, a pesquisa, o ensino, a divulgação e o culto a obra e vida de Rui Barbosa, para assim, contribuir para o conhecimento de diversidade cultural e para o fortalecimento da cidadania.

Segundo a Fundação Casa de Rui Barbosa (2018a), suas principais atividades são: manter, preservar e difundir o Museu Casa de Rui Barbosa, seu jardim histórico e o acervo bibliográfico e documental; oferecer qualificação e desenvolver pesquisas e publicações nas áreas de sua competência, além de utilizar seu espaço físico para atividades culturais.

- **Fundação Cultural Palmares - FCP**

Segundo a Fundação Cultural Palmares (2016) e Secretaria Especial da Cultura (2013), a Fundação foi criada em agosto de 1988, pela Lei. n. 7.668, sendo a primeira instituição pública voltada à promoção e preservação da arte e da cultura afro-brasileira. Dessa forma, preocupa-se com a igualdade racial e com a valorização das manifestações de matriz africana e implanta políticas públicas que potencializam a participação da população negra brasileira nos processos de desenvolvimento do País.

- **Fundação Nacional das Artes - Funarte**

De acordo com a Funarte (2010) e Secretaria Especial da Cultura (2013), a fundação foi criada em 1975, pela Lei n. 6.312, a Funarte tem como finalidade promover, estimular e desenvolver atividades culturais, incentivando à produção e à capacitação de artistas, o desenvolvimento da pesquisa, a preservação da memória e a formação de público para as artes no Brasil. Assim, é responsável pelo desenvolvimento de políticas públicas de fomento às artes visuais, à música, ao teatro, à dança e ao circo.

Segundo Funarte (2010), também concede bolsas e prêmios, mantém programas de circulação de artistas e bens culturais, promove oficinas, publica livros, recupera e disponibiliza acervos, provê consultoria técnica e apoia eventos culturais em todos os estados brasileiros e no exterior. Além de manter espaços culturais no Rio de Janeiro, São Paulo, Minas Gerais e Distrito Federal.

Conforme Funarte (2013, 2018), a fundação vem adaptando seu acervo, digitalizado itens que compõem sua vasta coleção – fotos, arquivos sonoros, textos, documentos – para serem disponibilizados ao público, na internet, tornando-o acessível a cada vez mais brasileiros. Esse processo se iniciou nos anos 2000, mas apenas em 2010, entrou no ar o Brasil Memória das Artes, totalmente dedicada aos acervos digitalizados da Funarte. Além deste, são responsáveis pelo Acervo Sergio Britto Digital e Acervo Funarte. Vale destacar o projeto Funarte Digital iniciado no ano de 2018 que, em parceria com o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), terá como um de seus objetivos dar maior acesso aos acervos da instituição e também utilizará a plataforma Tainacan para difusão dos acervos.

- **Fundação Biblioteca Nacional - FBN**

De acordo com a Biblioteca Nacional (2018) e Secretaria Especial da Cultura (2013c), o núcleo original da Biblioteca Nacional do Brasil teve início em 1808, como Real Biblioteca no Brasil, com a chegada de cerca de 60 mil peças, entre livros, manuscritos, mapas, trazidos por D. João VI e sua corte ao Rio de Janeiro. Em 1876, a instituição passa a se chamar definitivamente Biblioteca Nacional. Sua missão consiste na coleta, tratamento e conservação do patrimônio documental brasileiro em língua portuguesa e sobre o Brasil; bem como assegurar seu estudo, divulgação e as condições para sua disseminação.

Conforme Biblioteca Nacional (2018), Biblioteca Nacional Digital (2018) e Secretaria Especial da Cultura (2013), em 2006, criou-se a Biblioteca Nacional Digital - BNDigital, que integrou todas as coleções digitalizadas e vem sendo responsável pelos projetos da área. Em 2014 foi adquirido o software SophiA, de automação bibliográfica, e iniciada a migração dos catálogos para a plataforma, que é utilizada até hoje. Além da BNDigital, há disponível a Hemeroteca Digital Brasileira, um portal de periódicos nacionais que proporciona ampla consulta, pela internet, ao seu acervo de periódicos – jornais, revistas, anuários, boletins etc. – e de publicações seriadas.

2. Metodologia

2.1. Visão geral (Etapas)

O diagnóstico de potenciais acervos dos portais web das entidades vinculadas ao extinto Ministério da Cultura foi executado em 5 etapas:

1. Formulação do problema e identificação dos sites dos portais de cada uma das vinculadas
2. Identificação de pontos de acesso nos respectivos portais encontrados na primeira etapa
3. Descrição e categorização dos pontos de acesso encontrados na segunda etapa
4. Revisão dos resultados da descrição e categorização com o objetivo de garantir a normalização das categorias
5. Sintetização gráfica dos dados para apresentar os resultados do diagnóstico

A primeira etapa foi caracterizada pela formulação do problema de pesquisa, quais as formas, estratégias e modos de organização de itens culturais disponibilizados pelas instituições? Que orientou as demais etapas da pesquisa. A partir da definição do problema, foram identificados os links portais web das entidades vinculadas à Secretaria Especial da Cultura. Os links dos portais, foram obtidos em sua maioria do endereço web <http://antigo.cultura.gov.br/web/guest/entidades-vinculadas> que aponta qual o site de cada um dos portais web das entidades vinculadas à Secretaria Especial da Cultura. No caso do IBRAM (Instituto Brasileiro de Museus), foram considerados os links para os portais dos museus vinculados às entidades, identificados em <https://www.museus.gov.br/museus-ibram/>. Os links obtidos podem ser acompanhados na tabela 1 a seguir.

Tabela 1 - Links dos portais das entidades vinculadas ao MinC

Entidade Vinculada ao MinC	Link para o portal Web
ANCINE	https://www.ancine.gov.br/
IPHAN	http://portal.iphan.gov.br/
BN	https://www.bn.gov.br/
Fundação casa de Rui Barbosa	http://www.casaruibarbosa.gov.br/
FUNARTE	https://www.funarte.gov.br/
Fundação Cultural Palmares	http://www.palmares.gov.br/
IBRAM	https://www.museus.gov.br/museus-ibram/

A segunda etapa do processo de diagnóstico se deu a partir do acesso e identificação dos pontos de acesso aos conjuntos de itens presentes no portal de cada entidade vinculada. Esse trabalho foi executado manualmente, e consistiu na busca de páginas e sistemas de informação que caracterizassem um potencial acervo.

É importante destacar que nessa pesquisa foi considerado como acervo todo conjunto de itens encontrados nos portais das entidades vinculadas pesquisadas. Dessa forma, a pesquisa considerou além dos repositórios digitais (quando houvessem), conjuntos de itens sistematizados em páginas dispersas, considerando que potencialmente poderiam ser parte do acervo da entidade vinculada, e que em uma busca integrada seria interessante que houve a recuperação desses itens. Partiu-se do princípio que muitas instituições poderiam publicar itens culturais e, portanto, objetos de interesse público em formatos muito variados e distintos entre si, sendo muitos deles não necessariamente os mais eficientes em termos de estratégias de busca e recuperação da informação. Compreender esses formatos e as estratégias de publicação desses acervos de itens culturais tornaram-se um problema fundamental da pesquisa. Identificar o que potencialmente existe e possa ser melhorado em futuros projetos de reorganização da informação, constituição de repositórios digitais e aplicação de técnicas mais elaboradas de análise da informação que tornem a os objetos mais acessíveis pode se tornar uma contribuição significativa a área da cultura no país

A terceira etapa consistiu na descrição e categorização dos pontos de acesso encontrados. Esse processo de descrição foi realizado com o auxílio de uma tabela compartilhada on-line, que apresenta variáveis categóricas para cada ponto de acesso identificado, são elas: *tipo de estrutura do sistema de informação, ferramenta utilizada, licença, forma de extração dos dados, padrão de metadados, linguagens documentárias, regras de catalogação, estrutura do conjunto, tamanho do conjunto, e tipos de mídia*. Cada uma dessas categorias de análise é descrita na seção 3.2 deste relatório.

Esse processo de categorização foi realizado entre 03 e 17 de fevereiro de 2020 por dois pesquisadores doutorandos em Ciência da Informação, com temáticas de projeto próximas à temática desta pesquisa.

Na quarta etapa, os dados obtidos da categorização foram revisados e normalizados. Esse processo se deu com a revisitação de todos os links dos pontos de acesso das entidades vinculadas, verificando a adequação à cada categoria analítica identificada na terceira etapa, normalizando assim as categorias e garantindo a qualidade dos dados de descrição dos pontos de acesso das entidades vinculadas. Essa etapa foi realizada entre 27 de março e 02 de abril de 2020.

A quinta e última etapa marcou o desenvolvimento da síntese gráfica dos resultados e a produção deste relatório analítico do diagnóstico dos pontos de acesso das entidades vinculadas ao MinC. Nessa etapa foram calculados os números de pontos de acesso, bem como a frequência da ocorrência de cada categoria analítica que descreve esses pontos de acesso. Os resultados dessa etapa são apresentados na seção 4 deste relatório.

3. Variáveis da base de coleta

3.1. Variáveis de Identificação

As variáveis de identificação caracterizam os pontos de acesso recuperados dos conjuntos de itens nos portais das entidades vinculadas. Foram definidas três variáveis de identificação:

- **Provedor de Dados:** identifica o nome da entidade vinculada.

- **Fonte de Informação:** identifica o nome do ponto de acesso segundo o título da página web em que se encontra.
- **Link:** apresenta o link do ponto de acesso identificado.

3.2. Variáveis de categorização

As variáveis de categorização apresentam as tipologias em que cada conjunto de objeto digital identificado se encaixa de acordo com as categorias de análise elencadas na pesquisa. São elas:

- **Estrutura do sistema de informação:** a tipologia da estrutura do sistema de informação aqui proposta faz menção ao tipo de sistema de gestão e publicação de dados utilizados nos conjuntos de itens identificados no diagnóstico. Foram encontrados 5 tipos de sistema de informação:
 - **Página estática (HTML):** São conjuntos de itens expostos através de uma página HTML simples, onde foi identificado que o documento está integrado a página, tais como fotografias digitais exibidas na página web.
 - **Repositório Digital:** São conjuntos de itens expostos em um sistema específico para gestão, como DSpace, SophiA Web ou o Wordpress+Tainacan, que indica o uso do CMS WordPress com a plataforma de publicação de acervos Tainacan incorporada, por exemplo.
 - **Sistema de Gerenciamento de Conteúdo:** São conjuntos de itens expostos em páginas estruturadas a partir de um *Content Management System* (CMS), como Wordpress ou Drupal por exemplo.
 - **Documento:** São conjuntos de itens expostos em estrutura de arquivos no formato de listas, como lista de itens publicadas dentro de um arquivo em PDF, ou conjuntos de arquivos em ZIP, ou ainda expostos como pontos em um mapa, utilizando a ferramenta *Google My Maps*.
- **Ferramenta:** A categoria de ferramenta está diretamente ligada com a categoria de estrutura do sistema de informação que expõe os conjuntos de itens identificados no diagnóstico.

- **Da tipologia de Sistema de Gerenciamento de Conteúdo:** Drupal, Joomla e Wordpress.
- **Da tipologia de Repositório Digital:** DSpace, PHL, Pergamum, SophiA, Wordpress+Tainacan, OJS, OrtoDocs, Fotoweb 7, Google Arts and Culture e SICG.
- **Da tipologia de Documento:** Adobe
- **Da tipologia de Página Estática (HTML):** São mecanismos de estruturação da exposição dos pontos de acesso que não se enquadram propriamente na nomenclatura de software específico para gestão de informação, mas que faz uso de linguagem de marcação para publicação de informação em rede, como por exemplo o uso de PHP e HTML para criação de conteúdo em páginas da Web.
- **Não identificados:** São software acessíveis pelo portal web, que indicam expor conjuntos de itens, porém não contém nenhuma autoria ou identidade (<http://arquivo.bn.br/site/pages/bibliotecaDigital/passadomusical/script/index.asp>, http://arquivo.bn.br/portal/index.jsp?nu_pagina=75), ou ainda sistemas que não suportam mais o acesso (<http://portal.iphan.gov.br/pagina/detalhes/228>), e também as estruturas de arquivo no formato ZIP.
- **Licença:** Na categorização das licenças foram identificados 3 tipos de licenças utilizadas pelas entidades vinculadas para conceder acesso aos seus conjuntos de itens: **Copyright**, **Creative Commons**, e **CC BY-ND 3.0**. Ainda houveram casos em que não foi identificada explicitamente a licença de acesso e uso dos itens.
- **Forma de Extração de Dados:** Na categorização das formas de extração de dados, foram identificadas 7 categorias de obtenção das informações dos itens: **API**, **Harvester** (OAI-PHM), **CSV**, **KML**, **Pedido por E-mail**, **Download de documentos** (no caso de arquivos PDF e ZIP com a lista dos itens), **Raspagem de Dados** (processo automático ou semiautomático de se obter dados diretamente de sua página na Web).

- **Padrão de Metadados:** Na categorização dos padrões de metadados utilizados para catalogar os itens das entidades vinculadas foram identificadas 4 categorias: **Dublin Core**, [Dublin Core+](#) (quando o Dublin Core é estendido com mais metadados), **MARC**, **Não identificado** (quando entende-se que existe um padrão de metadados implícito no conjunto de itens, mas não foi reconhecido).
- **Linguagens Documentárias:** Na categorização das linguagens documentárias utilizadas para auxiliar na padronização da descrição dos itens, foram identificadas 4 categorias: **CDD**, [Base de Terminologia da FBN](#), [Vocabulário controlado sobre escravidão, abolição e pós-abolição](#) e **Não Identificado** (onde não foi identificada linguagem documentária de apoio explícita).
- **Regras de Catalogação:** Na categorização das regras de catalogação utilizadas para padronizar a forma de inserção de informações sobre os itens, foram identificadas 2 categorias: **AACR2**, e **Não Identificado** (onde não foram identificadas regras de catalogação explícitas).
- **Forma de visualização do Acervo:** Na categorização de forma como o acervo está publicado, que constitui da forma como o conjunto de itens é apresentado no portal web das entidades vinculadas, foram identificadas 3 categorias: **Coleções**, em que os itens estão distribuídos em coleções; **Exposição**, em que os itens estão organizados no formato de exposições (páginas HTML com parte do acervo por exemplo); **Hierárquica**, em que os itens estão organizados de forma hierárquica (como em listas de registros ou vídeos, ou ainda conjuntos de pastas por exemplo);
- **Quantidade de Itens:** O tamanho do acervo é uma variável descritiva que teve como objetivo mensurar ou estimar a quantidade de itens existente em cada ponto de acesso exposto no portal web de cada entidade vinculada. Alguns conjuntos de itens não tiveram tamanhos identificados pela dificuldade de encontrar essa informação nos próprios sistemas de publicação, como o SophiA Web, que em determinados pontos de acesso não permite fazer uma busca por todo o acervo,

não informando assim em nenhuma página pública o total de itens indexados pelo sistema.

- **Mídia:** Na categorização de mídias, que apontam como os itens podem ser exportados, foram identificadas 8 categorias: **E-book**; **PDF**; **PDF/A**; **Áudio**; **Imagem**; **Vídeo**; e **Texto**, quando somente os registros foram identificados e o item cultural não foi digitalizado e disponibilizado on-line.

4. Resultados

Os resultados sintetizados a partir do cálculo de ocorrência das tipologias das variáveis categóricas podem ser acompanhados graficamente nos próximos tópicos.¹

4.1. Números Totais

Esse tópico apresenta de forma geral os resultados dos conjuntos de itens encontrados nos pontos de acesso identificados para cada entidade vinculada ao extinto Ministério da Cultura.

No total foram encontrados **217 pontos de acesso** referentes à potenciais acervos nos portais web das vinculadas, que representam um total de **2.537.921 itens**. Ainda em 28 pontos de acesso não foi possível mensurar a quantidade de itens por falta de recurso técnico disponível para acesso à informação. É importante ressaltar, portanto, que o número de itens disponíveis nos sites pode ser maior do que o mencionado acima.

Ao observar a proporção de pontos de acesso por vinculada (figura 1) o IPHAN demonstra maior evidência (52,56%) em relação às demais entidades com mais da metade dos pontos encontrados. Já a Biblioteca Nacional, o IBRAM e a Fundação Casa de Rui Barbosa demonstram entre 10% e 18% dos pontos encontrados. Com menos expressividade a FUNARTE, Ancine e Fundação Palmares apresentam uma proporção de pontos de acesso menor que 5% dentre todas as vinculadas.

¹ É importante ressaltar que 3 pontos de acesso foram identificados, mas o acesso ao sistema não foi efetivado devido problemas no próprio sistema que impedem a busca ou exploração por objetos. São 2 pontos de acesso vinculados ao IPHAN (<http://portal.iphan.gov.br/pagina/detalhes/228>, <http://portal.iphan.gov.br/pagina/detalhes/219/>) além de um ponto de acesso da Biblioteca Nacional (<http://arquivo.bn.br/site/pages/bibliotecaDigital/passadomusical/script/index.asp>). Dessa forma não foram analisados dados sobre esses pontos de acesso.

Esses resultados refletem a característica de granularidade do acervo das instituições, já que quanto maior a quantidade de pontos de acesso maior o esforço do usuário para navegar no portal para acessá-los e menor sua capacidade de visualizá-los de forma integrada. O mesmo fator dificulta processos de integração dos conjuntos de itens dos acervos, que deve então considerar recuperar informação de diferentes pontos de acesso, com padrões de extração de dados diferentes.

Pontos de Acesso por Vinculada

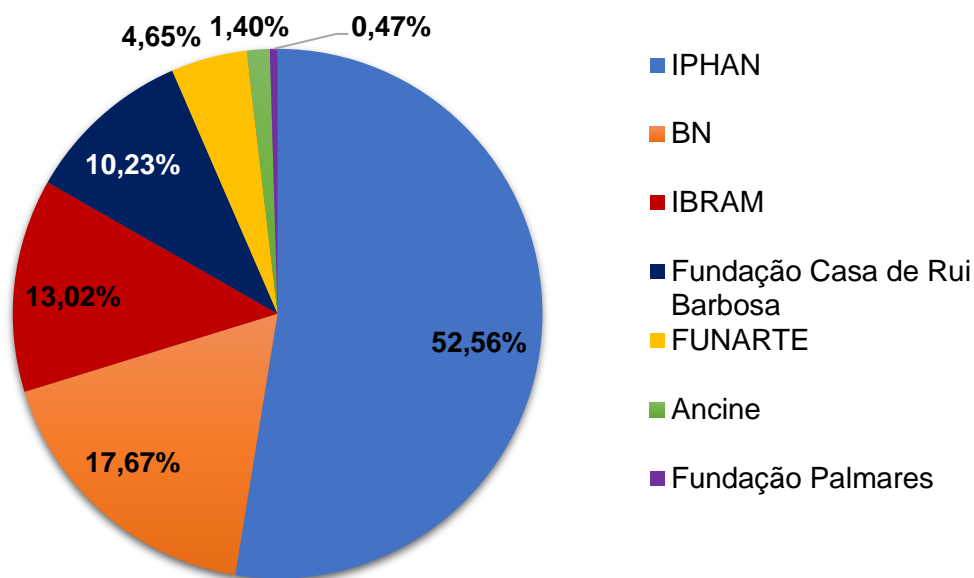


Figura 1 – Pontos de acesso por vinculada

Já observando o total de itens por entidade vinculada (figura 2), a Biblioteca Nacional apresenta maior evidência com mais de dois milhões de itens, quase 6 vezes mais itens que todas as outras vinculadas juntas. O IPHAN (153.833) e a FUNARTE (145.359) representam as duas entidades seguintes com mais de 100 mil itens cada uma. Já o IBRAM, a Fundação Casa de Rui Barbosa e a Ancine têm entre 6 mil e 43 mil itens. E com a menor quantidade de itens identificada, a Fundação Palmares com 9 Itens.

É importante ressaltar que essa quantidade de itens encontrada pode variar, à medida que existam pontos de acesso não encontrados por essa pesquisa, além dos pontos de acesso que dificultam a descoberta do total de itens mencionados no segundo parágrafo do tópico 4.

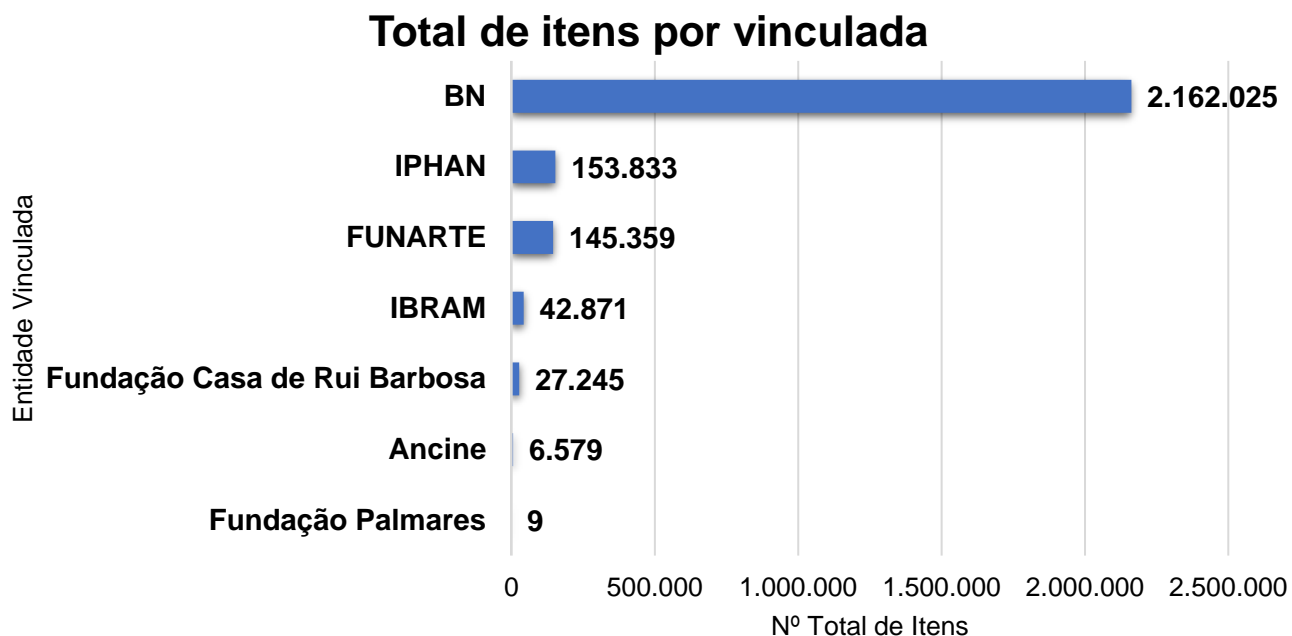


Figura 2 - Total de itens por vinculada

4.2. Pontos de acesso encontrados por variável categórica

Esse tópico apresenta a quantidade de pontos de acesso identificados que se enquadram em cada tipologia de cada variável categórica. Permitindo entender quais as categorias mais representativas no contexto de cada variável analítica utilizada.

Quanto ao tipo estrutura do sistema de informação utilizado para a divulgação dos itens nos pontos de acesso (figura 3), é evidente a expressividade do uso de páginas estáticas em HTML e repositórios digitais para a disponibilização do objeto on-line. O uso de páginas HTML reflete um uso limitado de recursos tecnológicos de sistematização dos itens, que poderiam ser armazenados em repositórios digitais com funcionalidades específicas.

Quantidade de Pontos de Acesso por Estrutura do Sistema de Informação

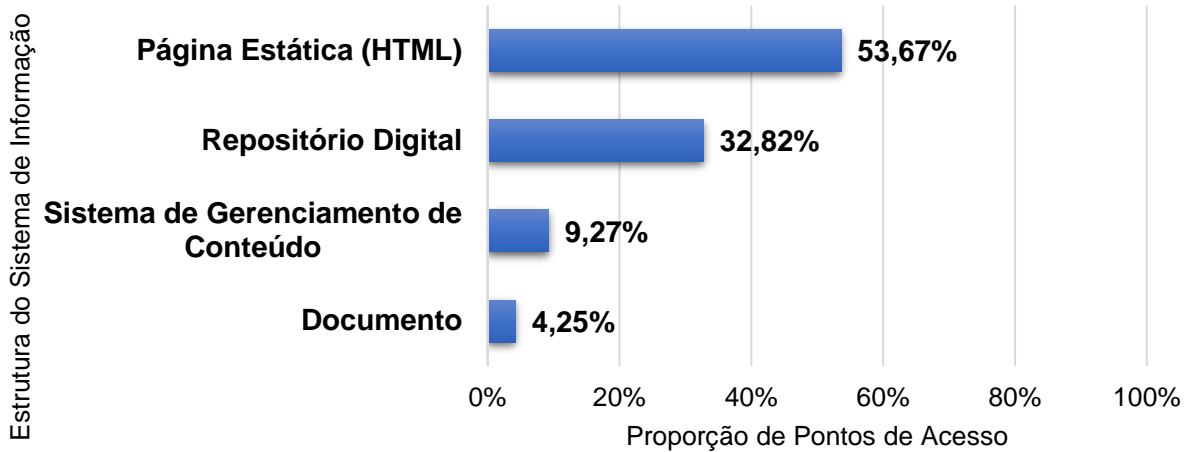


Figura 3 – Proporção da quantidade de pontos de acesso por estrutura dos sistemas de informação

Ao observar as ferramentas utilizadas para a disposição dos itens nos pontos de acesso (figura 4), a categoria página estática (HTML) é mais evidente dentre as outras tipologias, as demais se subdividiram em mais de um tipo de ferramenta/software.

Quantidade de Pontos de Acesso por Ferramenta Utilizada

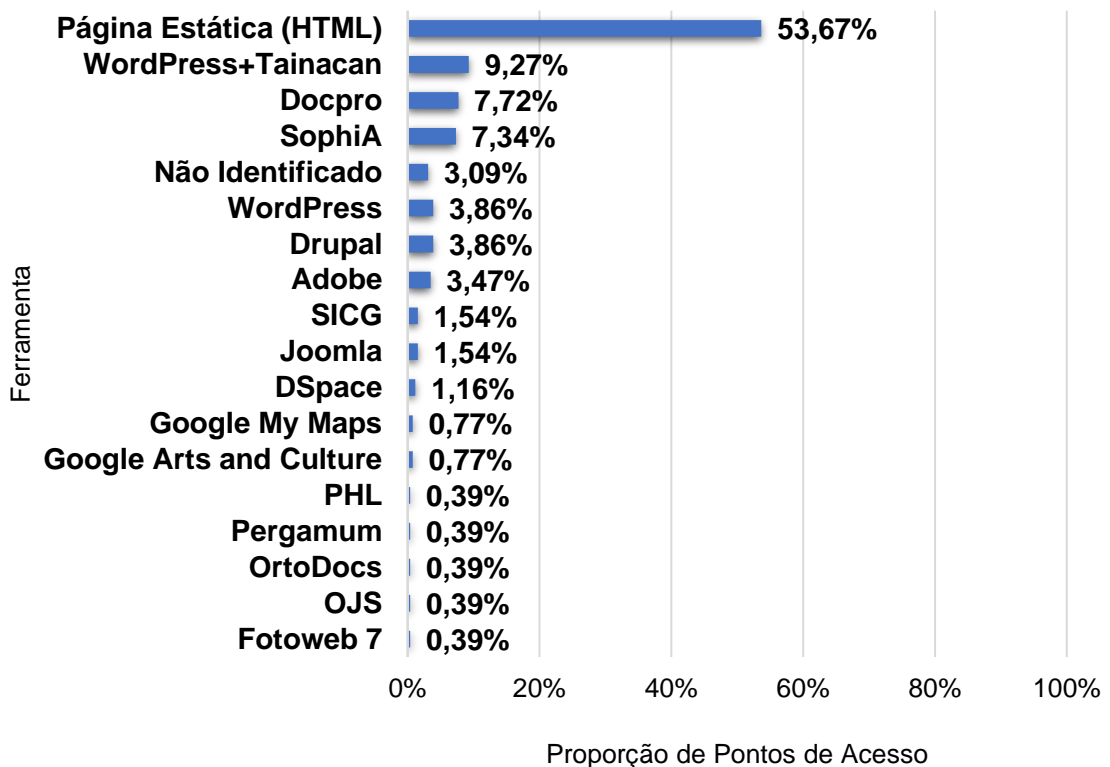


Figura 4 – Proporção de pontos de acesso por ferramenta utilizada

Quanto ao tipo de licença de uso definida para os itens (figura 5), essa pesquisa não conseguiu identificar nos pontos de acesso dos itens as licenças vigentes para uso do conteúdo. Esse resultado dá abertura para o questionamento de como as entidades vinculadas explicitam a licença dos seus acervos através de seus portais web, já que nas interfaces de acesso aos itens essa informação em sua maioria não foi encontrada.

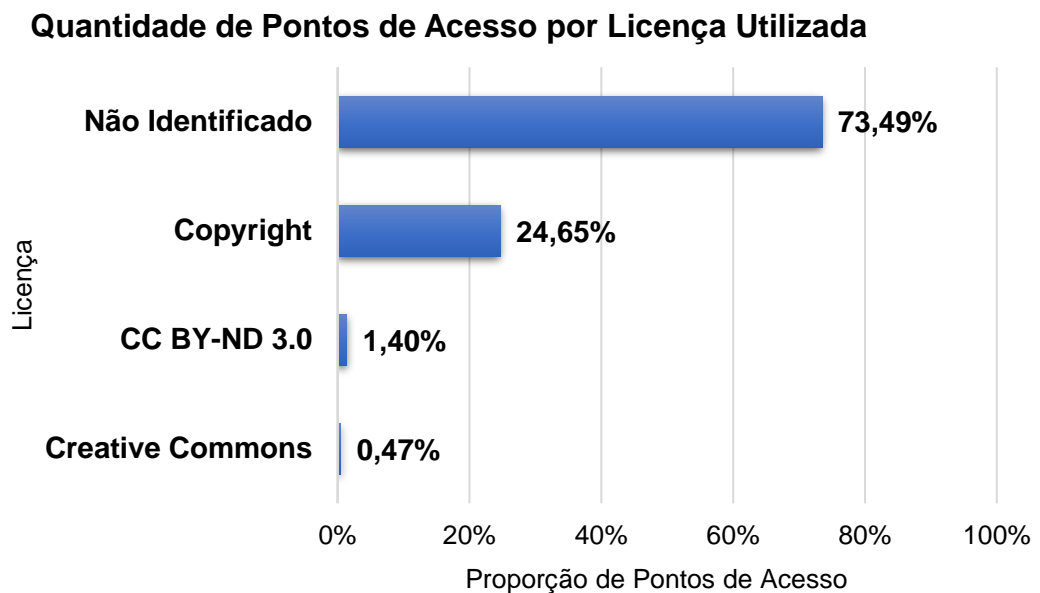


Figura 5 - Proporção de pontos de acesso por licença

Ao observar os padrões de metadados utilizados para descrever os itens (figura 6), novamente na maioria dos casos (90,23%) o padrão de metadados não é explícito ou não foi possível identificar um padrão pela simples observação dos dados.

Quantidade de Pontos de Acesso por Padrão de Metadados Utilizado

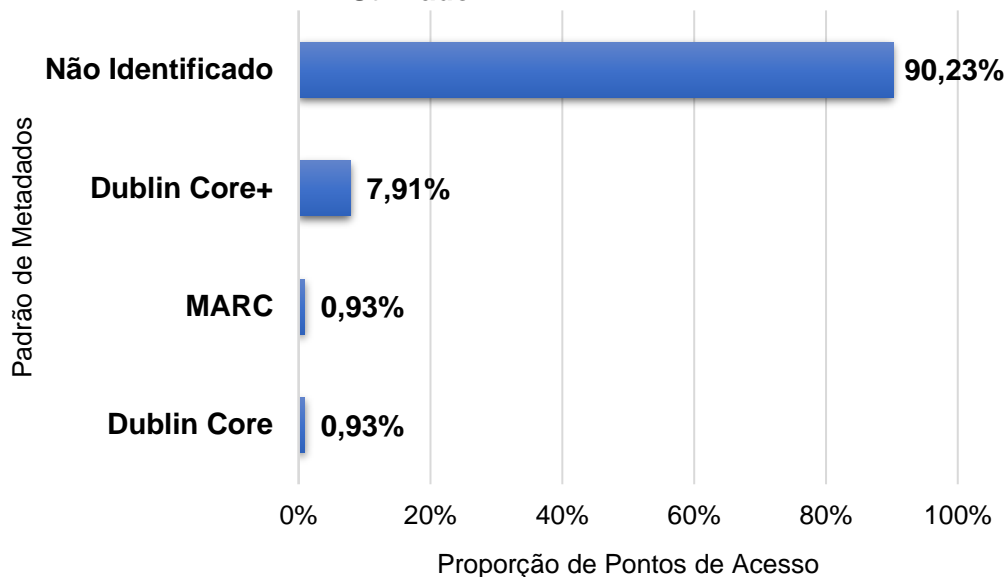


Figura 6 - Proporção de pontos de acesso por padrão de metadados utilizado

Novamente, agora no caso das linguagens documentárias utilizadas (figura 7), os resultados apontam para a maioria (97,70%) dos pontos de acesso sem linguagem documentária explícita.

Quantidade de Pontos de Acesso por Linguagem Documentária Utilizada

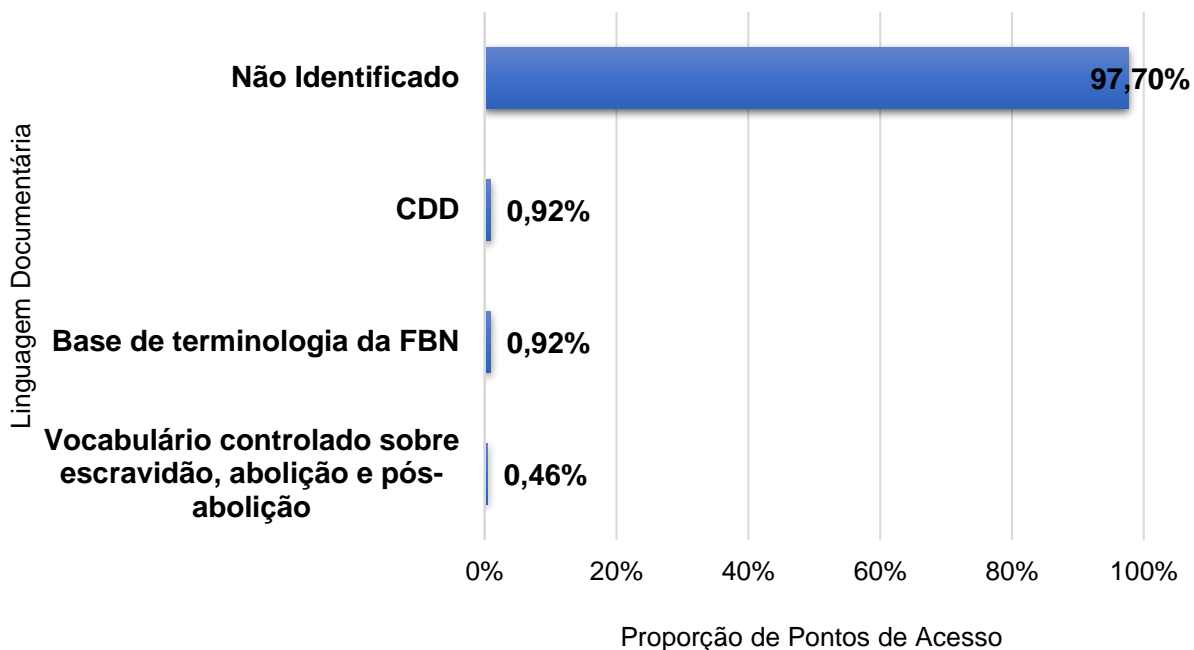


Figura 7 - Proporção de pontos de acesso por linguagens documentárias utilizadas

Ainda, analisando a variável de regras de catalogação (figura 8), também apresenta a característica comum da documentação sobre os pontos de acesso, em 99,07% não foram identificadas quais as regras de catalogação aplicadas na indexação dos itens.

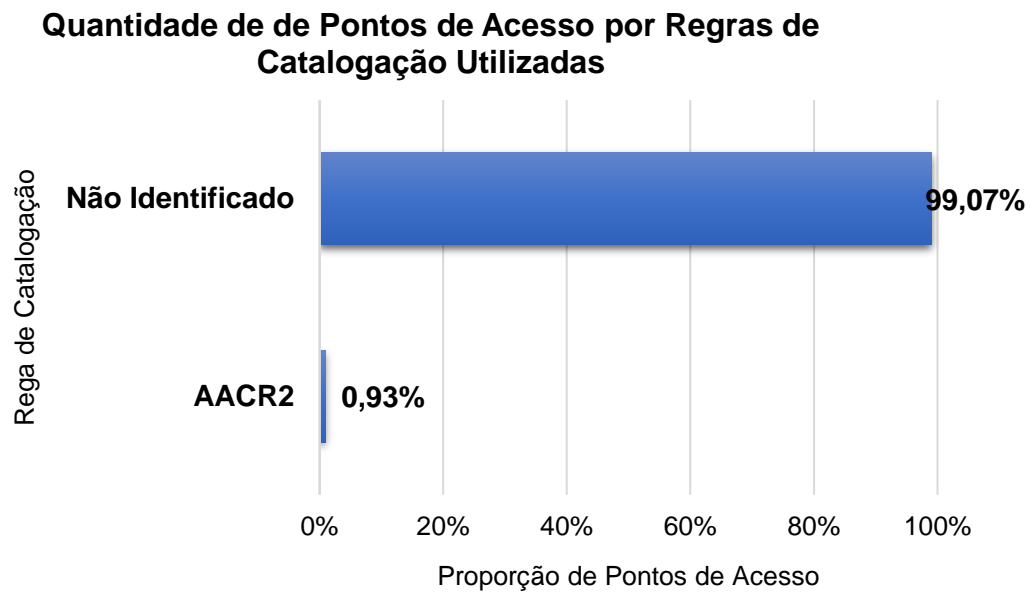


Figura 8 - Proporção de pontos de acesso por regras de catalogação utilizadas

Quanto à forma como os pontos de acesso são visualizados nos portais web das entidades vinculadas (figura 9), em sua maioria (76,74%) estão dispostos em formato hierárquico, uma forma mais simples de exibição e que não garante necessariamente interconexão entre os itens.

Quantidade de Pontos de Acesso por Forma de Visualização do Acervo

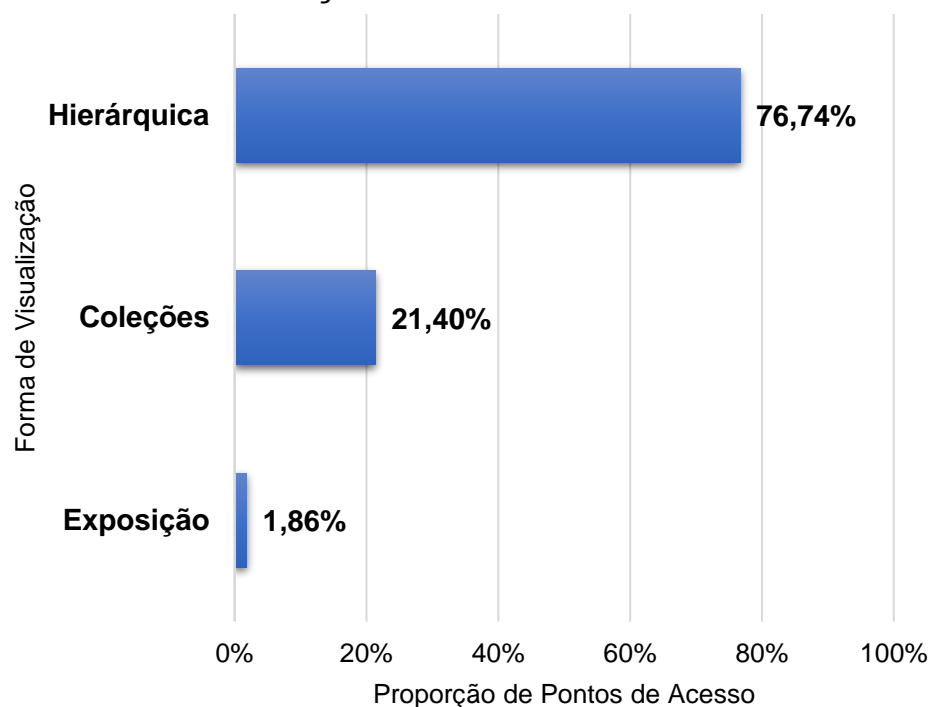


Figura 9 - Proporção de pontos de acesso por forma de visualização dos pontos de acesso

Ao analisar as formas como os pontos de acesso podem ser extraídos (figura 10), para a maior parte (69,81%) dos pontos é necessário aplicar técnicas de raspagem de dados, o que reflete na dificuldade de obter os itens para reuso já que são técnicas de complexidade mais alta, comparadas com as exportações por CSV ou API. Tal resultado demonstra uma enorme fragilidade na constituição das estratégias tecnológicas dos acervos culturais brasileiros, tornando o desafio de construir ferramentas digitais de busca integrada desses acervos bastante mais complexa.

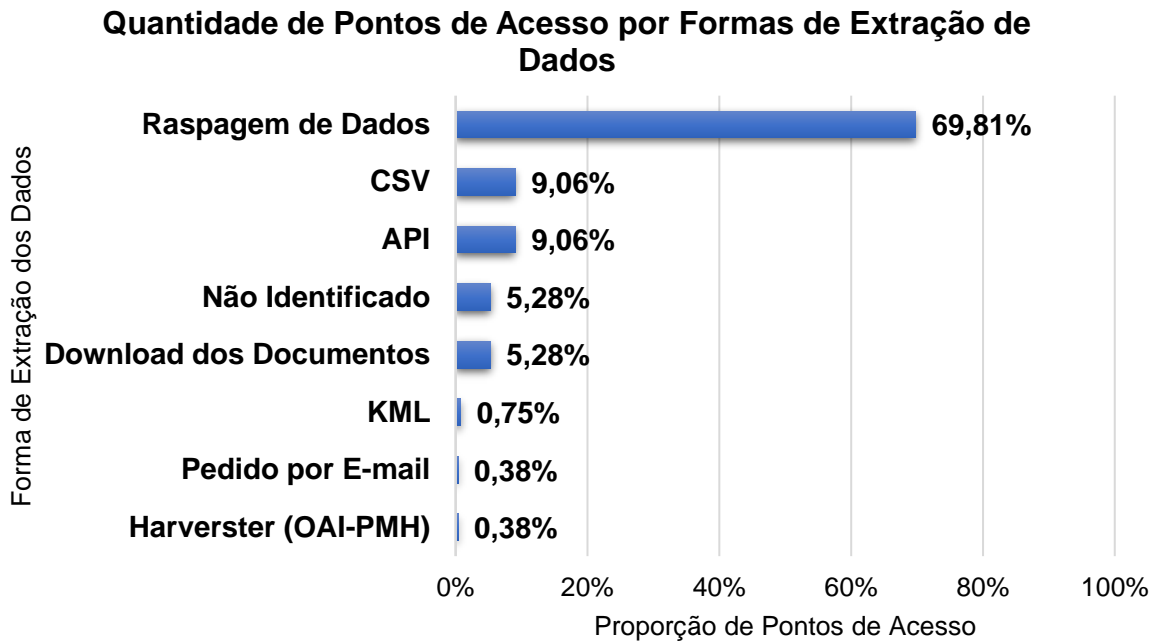


Figura 10 - Proporção de pontos de acesso por formas de extração dos dados

Já quanto as mídias disponíveis nos pontos de acesso (figura 11), houve uma proporção maior em itens no formato de imagem (33,48%) e no formato texto (25,75%) em que só os registros existem.

Quantidade de Pontos de Acesso por Mídias Disponíveis no Acervo

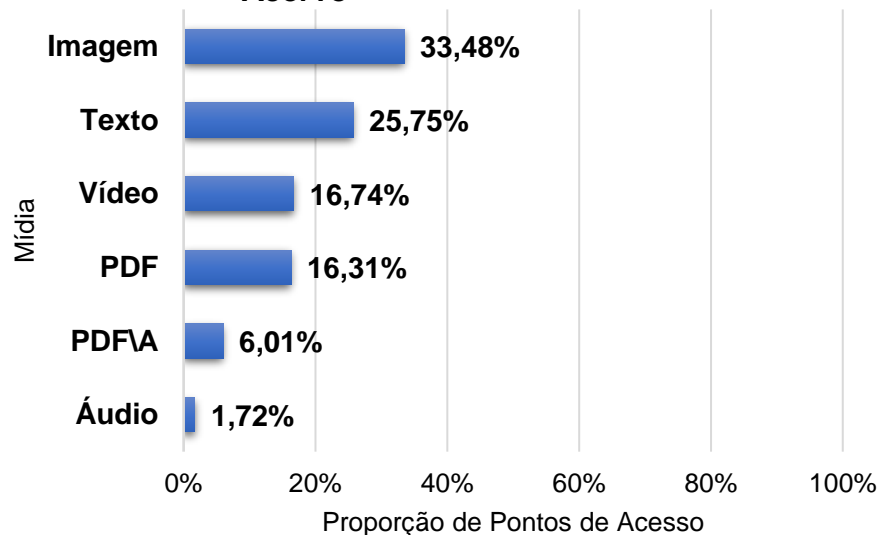


Figura 11 - Proporção de pontos de acesso por mídias disponibilizadas

4.3. Descrição por vinculadas

Neste tópico serão apresentados os resultados dos tipos de pontos de acesso para cada variável analítica em relação às entidades vinculadas à secretaria especial de cultura. O objetivo dessas visualizações é permitir observar as variáveis distribuídas por vinculada e comparar os valores.

Ao analisar a quantidade de pontos de acesso quanto à tipologia da estrutura dos sistemas de informação por vinculada (figura 12), observa-se o uso de páginas estáticas (HTML) é mais concentrado no IPHAN, e com menor proporção na FCRB. Como já mencionado, o uso de páginas HTML para disponibilizar os itens têm limitações que podem interferir no seu acesso e reuso, já que o conteúdo fica disperso pelo site e não concentrado em um repositório que permite o gerenciamento dos itens.

Outra observação interessante é como o uso de repositórios digitais é bem distribuído entre as vinculadas, mostrando que esse tipo de sistema é mais comum entre as vinculadas, com exceção da Ancine e da Fundação Palmares em que não foi identificada a presença de repositórios digitais.

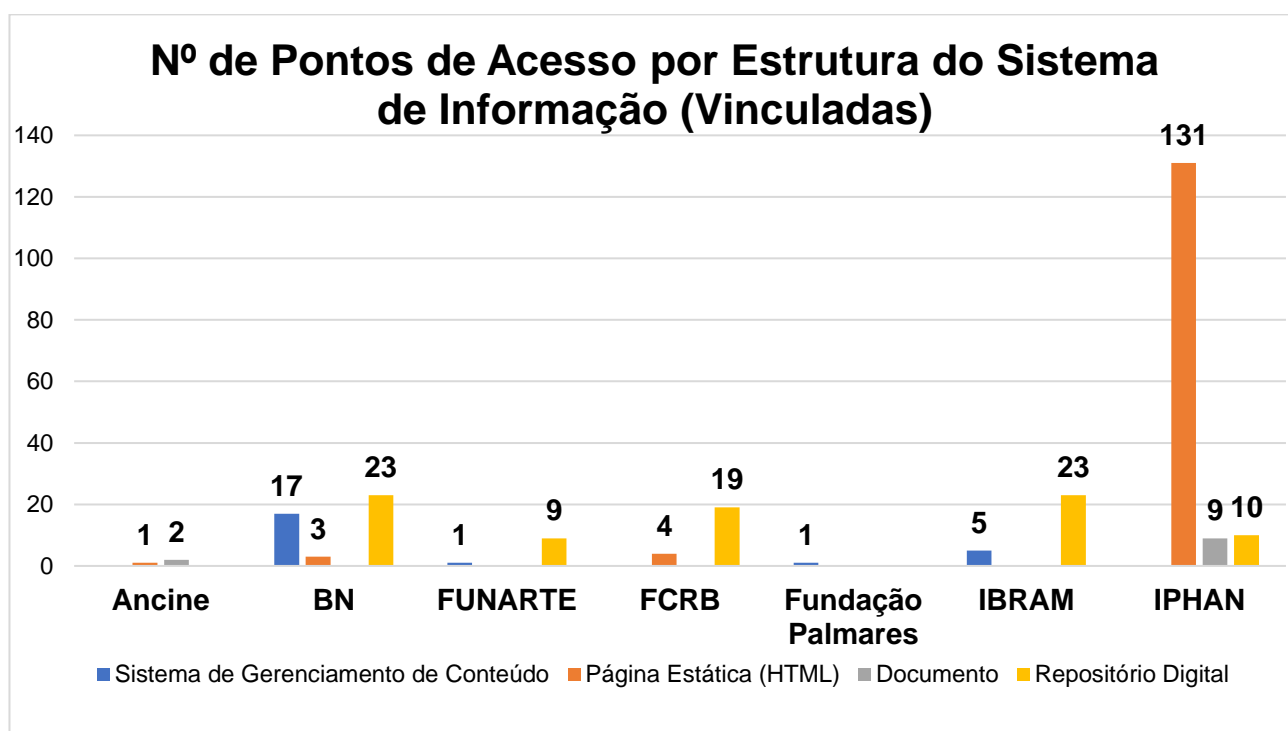


Figura 12 - Nº de Pontos de Acesso por Estrutura do Sistema de Informação (Vinculadas)

Quanto à ferramenta empregada nas soluções de disponibilização do acervo on-line (figura 13), novamente observa-se uma discrepância no uso de páginas estáticas em HTML pelo IPHAN, seguindo os mesmos preceitos avaliados no gráfico anterior.

A distribuição do uso de softwares ainda chama atenção pela característica de que cada vinculada adota um tipo específico de software para disponibilizar seu acervo on-line, indicando que não é um uso comum entre elas, mas sim um uso específico de cada uma. Como no caso da Funarte e do IBRAM o uso do Tainacan e do WordPress em conjunto, na Biblioteca Nacional o uso do Sophia Web e do Drupal, o uso do Docpro na FCRB e com menor expressão o uso do Google My Maps na Ancine e do WordPress na Fundação Palmares.

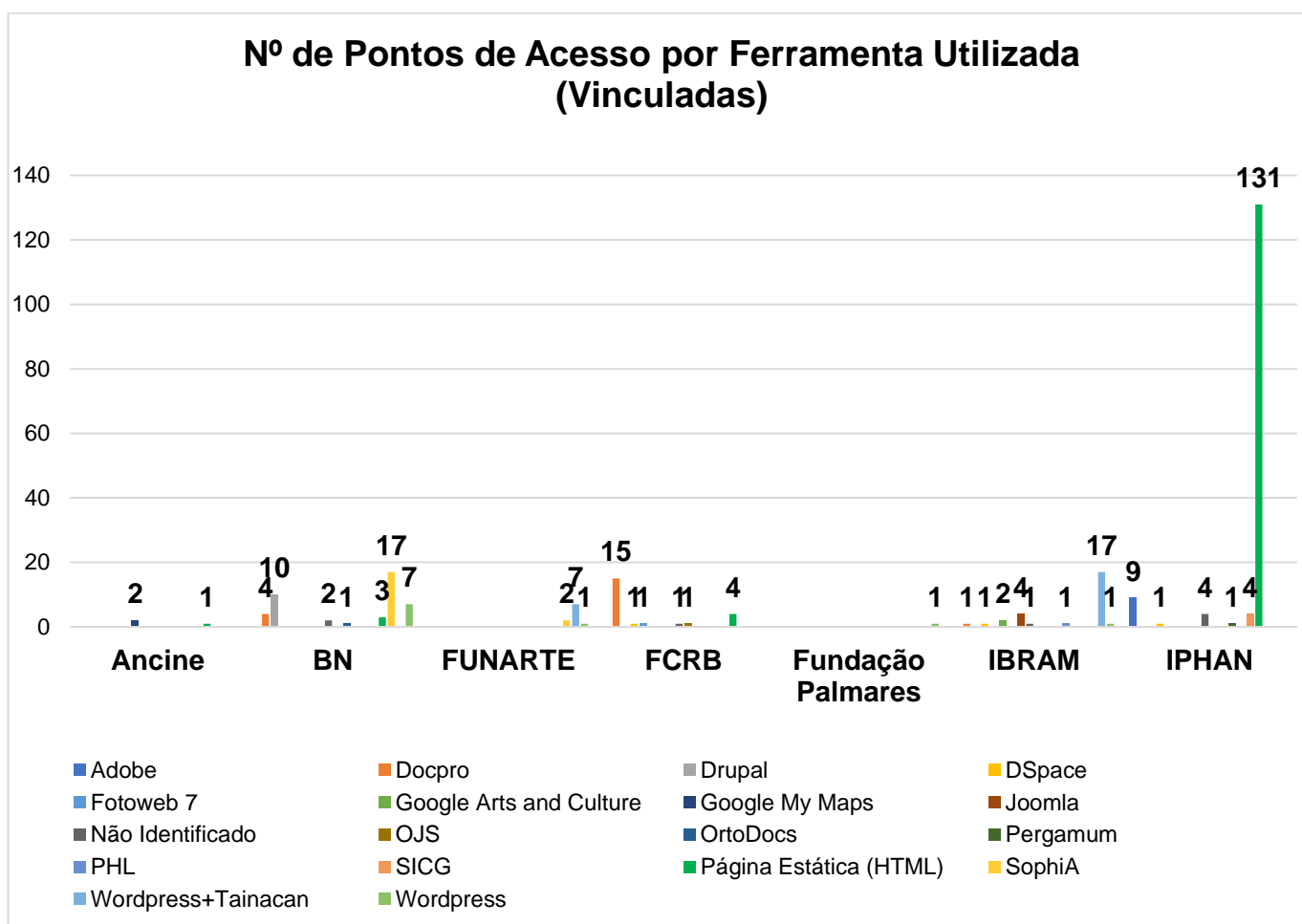


Figura 13 - Nº de Pontos de Acesso por Ferramenta Utilizada (Vinculadas)

Ao observar a documentação dos pontos de acesso, quanto à licença de uso (figura 14), apesar de a condição de não identificação da licença aparecer com maior expressão no IPHAN, é comum em todas as vinculadas, o que mostra a dificuldade de acessar essa informação através do portal web.

Em alguns pontos de acesso existe de maneira explícita a licença em vigor, como no caso da BN e da FCRB com o Copyright e Creative Commons, no caso do IBRAM ainda aparece o uso do CC BY-ND 3.0 e também do Copyright.

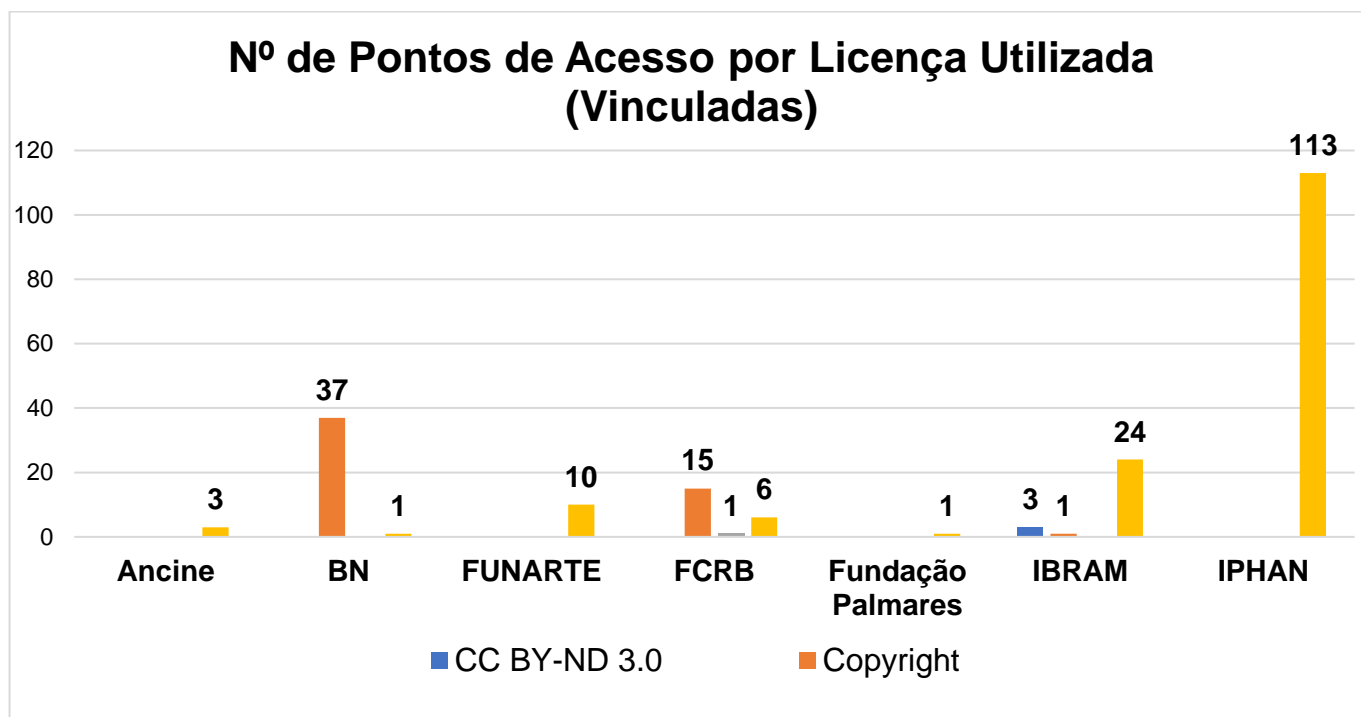


Figura 14 - Nº de Pontos de Acesso por Licença Utilizada (Vinculadas)

Continuando no contexto da documentação dos pontos de acesso, quanto aos padrões de metadados utilizados na descrição dos itens (figura 15), novamente em sua maioria não foram identificados de maneira explícita os padrões de metadados utilizados pelas entidades vinculadas. À exceção do Dublin Core que é apresentado em alguns pontos de acesso da FCRB, Fundação Palmares e do IPHAN, no caso da FUNARTE e da BN existe a variação do Dublin Core com alguns metadados específicos adicionados, e ainda é observado o uso do MARC no IPHAN e na FUNARTE.

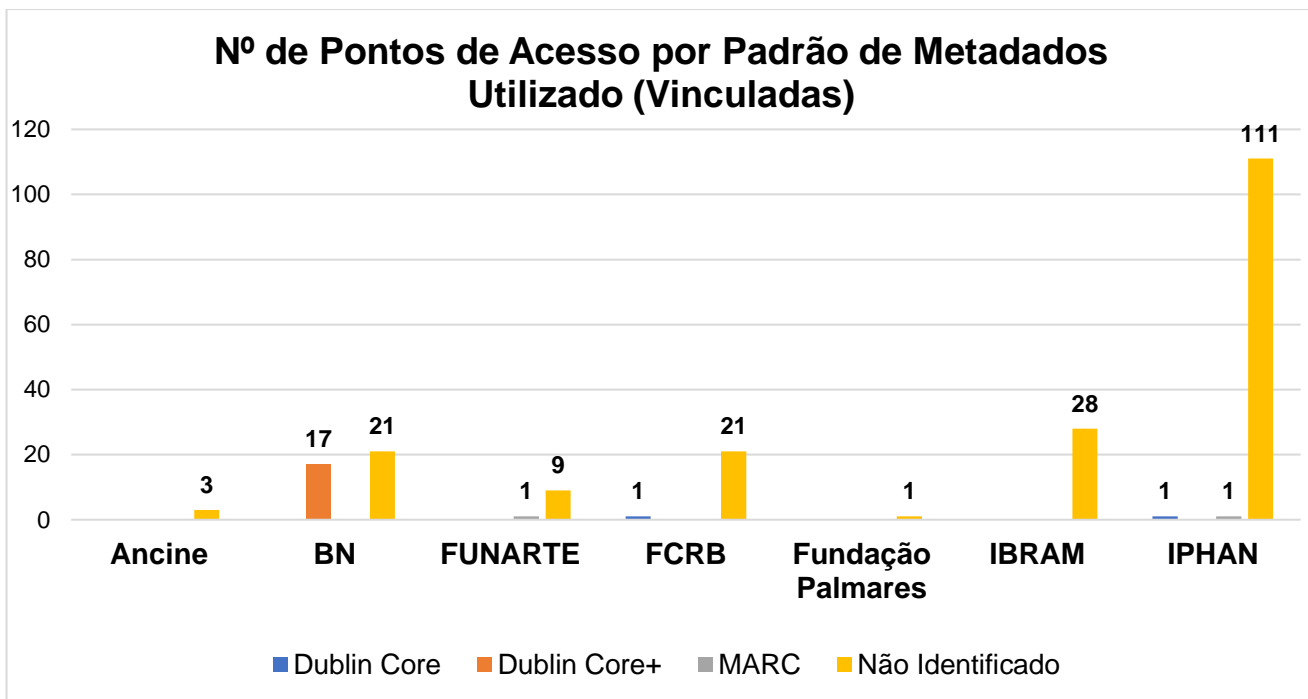


Figura 15 - Nº de Pontos de Acesso por Padrão de Metadados Utilizado (Vinculadas)

Já quanto às linguagens documentárias referentes aos pontos de acesso (figura 16), novamente não foi identificado nenhum tipo de linguagem documentária na maioria dos pontos. No caso da biblioteca nacional em alguns pontos de acesso foram explicitadas a base terminológica da FBN e o CDD, bem como na FCRB o uso de um vocabulário controlado específico.

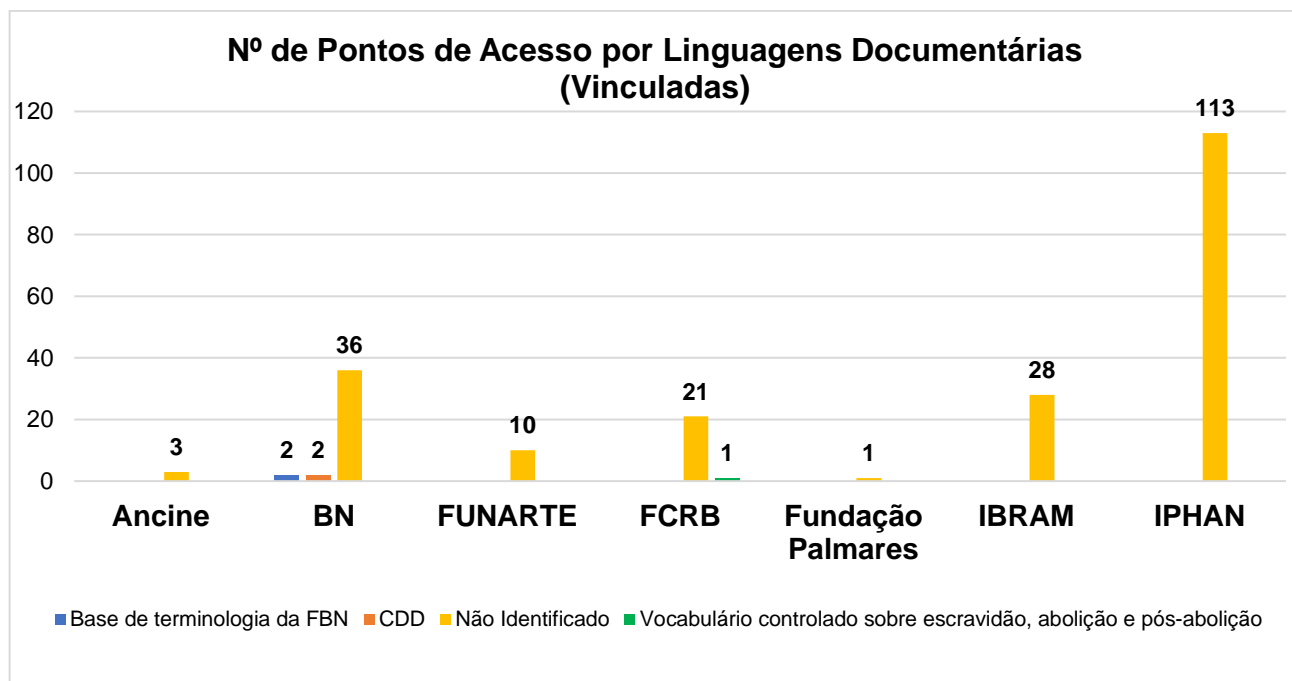


Figura 16 - Nº de Pontos de Acesso por Linguagens Documentárias (Vinculadas)

Ao observar as regras de catalogação aplicadas aos pontos de acesso (figura 17), novamente a proporção de regras não identificadas se distribui na maioria dos pontos de acesso recuperados das entidades vinculadas. Somente na biblioteca nacional em poucos pontos foi explícito o uso de regras de catalogação através da documentação do AACR2.

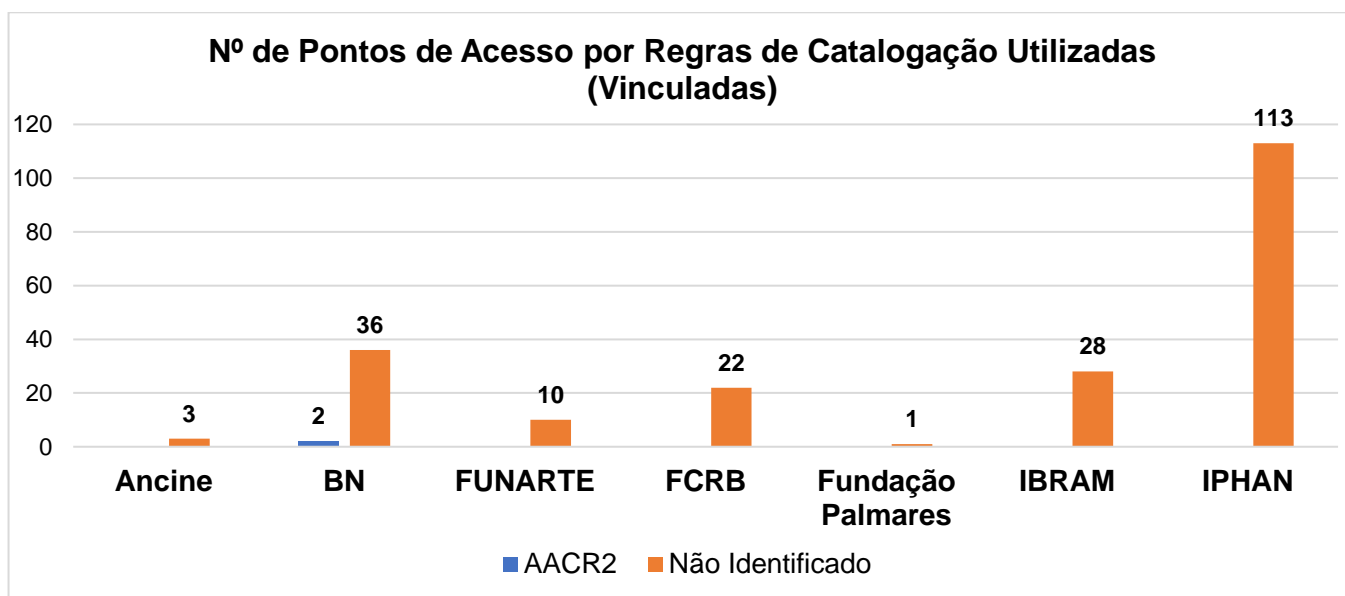


Figura 17 - Nº de Pontos de Acesso por Regras de Catalogação Utilizadas (Vinculadas)

Quanto à forma como os acervos podem ser recuperados (figura 18), a técnica de raspagem de dados foi o meio mais identificado para se obter os dados dos pontos de acesso das entidades vinculadas, resultado que é derivado da forma como as entidades expõem seus itens no portal, a falta de um sistema que permita exportação dos itens de maneira aberta e estrutura leva a necessidade de usar uma técnica mais complexa como a raspagem de dados.

Chama atenção ainda no IBRAM e na FUNARTE as possibilidades na maioria os pontos de acesso de obtê-los por API ou CSV que são meios mais estruturados e de menor complexidade, reflexo do uso do Tainacan + WordPress como repositório para seus itens.

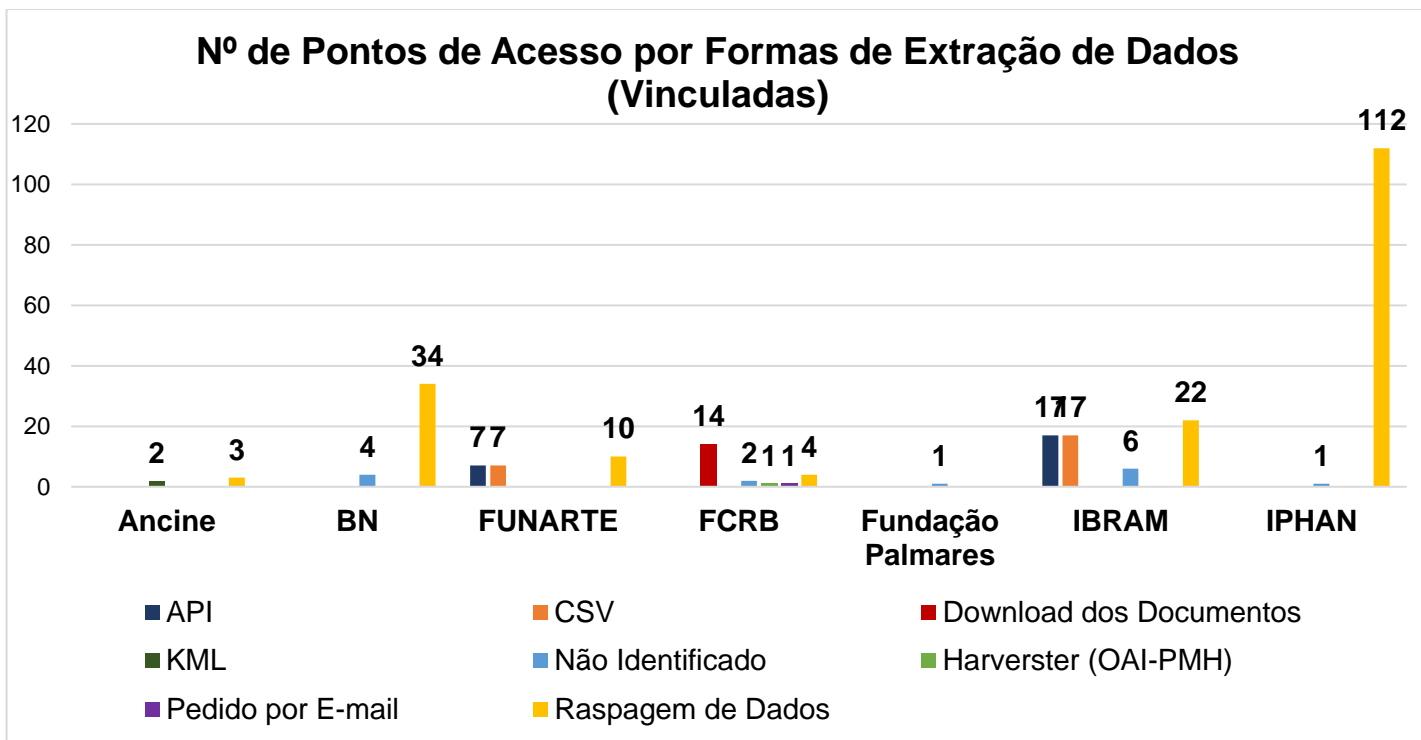


Figura 18 - Nº de Pontos de Acesso por Formas de Extração de Dados (Vinculadas)

Quanto à forma de visualização do acervo (figura 19), a forma de exposição hierárquica é comum aos pontos de acesso de todas as vinculadas, ainda chama atenção a forma de exposição em coleções, característica das funcionalidades de repositórios digitais que aparece nos portais web da BN, FUNARTE, FCRB e IBRAM.

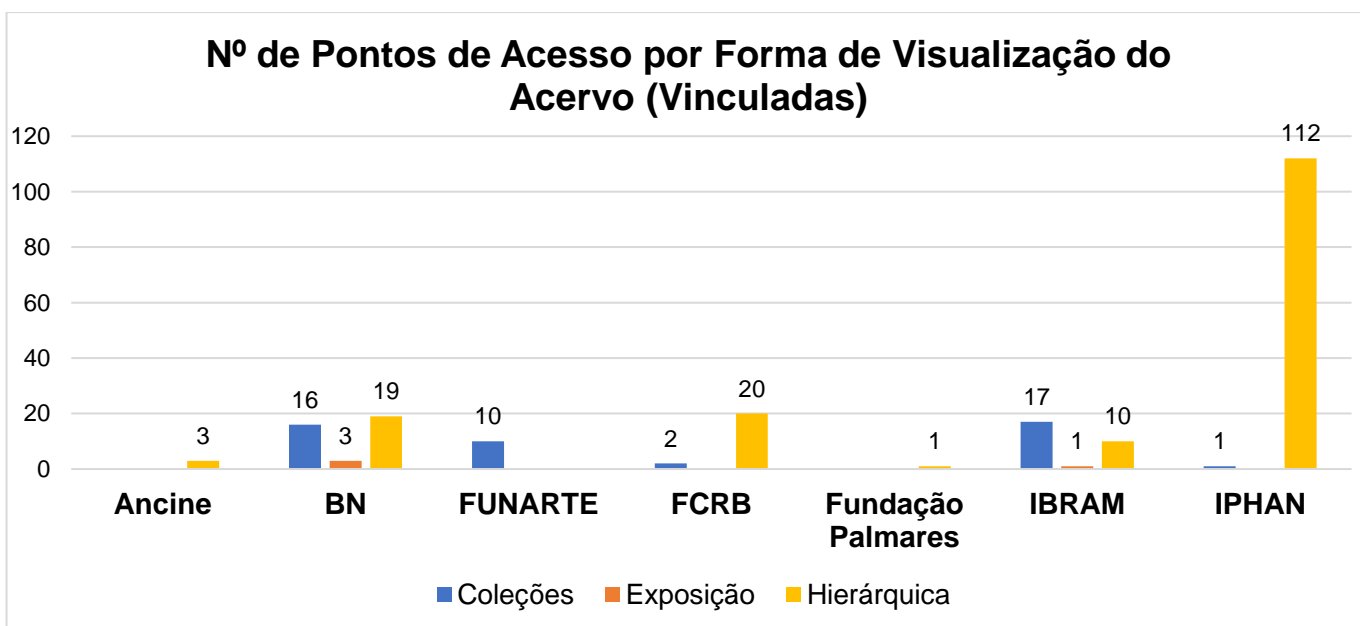


Figura 19 - Nº de Pontos de Acesso por Forma de Visualização do Acervo (Vinculadas)

Ao observar os tipos de mídia em que os itens são disponibilizados on-line (figura 20), é mais comum encontrar itens disponíveis no formato de PDF e imagem entre as vinculadas. Porém, nota-se a variabilidade de mídias do IPHAN, da FUNARTE e da BN, que disponibilizam itens em quase todos os formatos identificados.

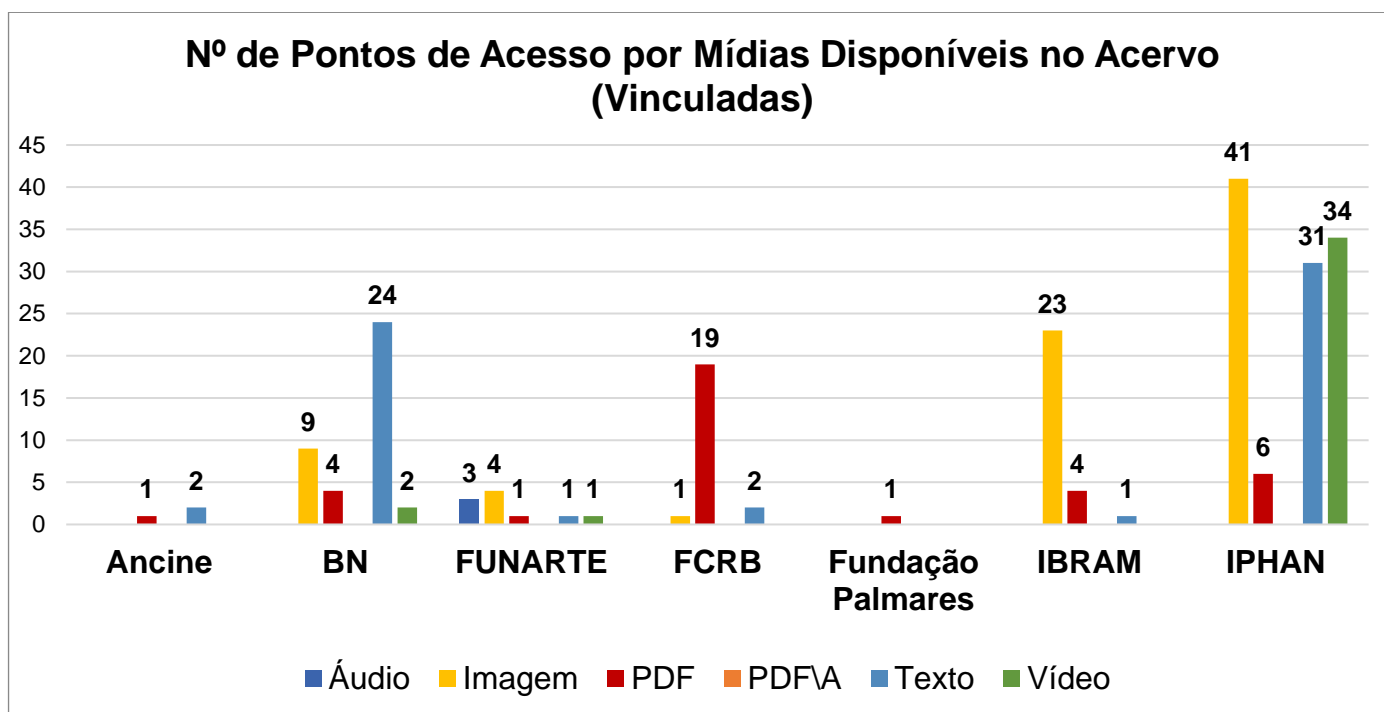


Figura 20 - Nº de Pontos de Acesso por Mídias Disponíveis no Acervo (Vinculadas)

5. Considerações Finais

Entende-se que este relatório satisfaz a realização do primeiro produto “Mapeamento sistemático das formas de organização da informação dos repositórios digitais de cultura” do projeto FAPESP “Interoperabilidade entre os repositórios digitais do patrimônio cultural brasileiro: da web semântica e dados abertos ligados às ferramentas de busca e recuperação da informação”.

De forma geral os resultados encontrados apontam algumas questões importantes:

- dificuldade de se encontrar documentação sobre os acervos das vinculadas. Como apresentam os resultados analisados das variáveis de *Linguagens Documentárias*, *Regras de Catalogação*, *Padrão de Metadados* e *Licenças*, o que significa um potencial comprometimento da interpretação da organização

da informação desses acervos. Para agregar os dados é fundamental o esclarecimento da documentação dos pontos de acesso, para entender a correspondência conceitual dos metadados utilizados entre as entidades vinculadas por exemplo.

- grande quantidade de itens sem condição de coleta formal, ou seja, poucos pontos de acesso permitem a coleta de dados através de aplicações como API ou um Harvester (OAI-PMH) ou ainda formatos abertos de dados estruturados como planilhas em CSV. Isso culmina na necessidade de coleta desses dados de maneira mais complexa, como raspagem de dados (Web Scraping) que consistem em identificar como os dados estão estruturados através de suas páginas web e desenvolver scripts de coleta para cada modo de disponibilização dos objetos.

Dessa forma, os resultados obtidos da coleta analítica dos pontos de acesso presentes nos portais Web das entidades vinculadas à Secretaria Especial da Cultura apresentam um diagnóstico suficiente para subsidiar as próximas etapas do projeto, que preveem a coleta e agregação dos dados identificados neste relatório.

6. Referências

ANCINE. Apresentação. Brasil: Agência Nacional do Cinema - Secretaria Especial da Cultura - Governo Federal. 2018. Disponível em: <<https://www.ancine.gov.br/pt-br/ancine/apresentacao>>. Acesso em 25 maio 2020.

BAGGIO, Claudia Carmem; FLORES, Daniel. Documentos digitais: preservação e estratégias. Biblos, Rio Grande, v. 27, n. 1, p. 11-24, 2013.

BIBLIOTECA NACIONAL DIGITAL. Hemeroteca Digital Brasileira. Brasil: Fundação Biblioteca Nacional – Secretaria Especial da Cultura – Governo Federal. 2018. Disponível em: <<http://bndigital.bn.gov.br/hemeroteca-digital/>>. Acesso em 25 maio 2020.

BIBLIOTECA NACIONAL. Histórico. Brasil: Fundação Biblioteca Nacional – Secretaria Especial da Cultura – Governo Federal. 2018. Disponível em: <<https://www.bn.gov.br/sobre-bn/historico>>. Acesso em 25 maio 2020.

COMITÊ GESTOR DA INTERNET NO BRASIL. Tic cultura 2018 = Survey on the use of information and communication technologies in brazilian cultural facilities : Ict in culture 2018 [livro eletrônico] / Núcleo de Informação e Coordenação do Ponto BR. -- São Paulo, 2019.

FUNARTE. Brasil Memória das Artes. Brasil: Associação Cultural Funarte – Fundação Nacional de Artes – Secretaria Especial da Cultura – Governo Federal. 2013. Disponível em: <<http://www.funarte.gov.br/brasilmemoriadasartes/acervo/o-projeto>>. Acesso em 25

FUNARTE. Funarte Digital: novo portal a caminho da implementação. Brasil: Associação Cultural Funarte – Fundação Nacional de Artes – Secretaria Especial da Cultura – Governo Federal. 2018. Disponível em: <<http://www.funarte.gov.br/artes-integradas/funarte-digital-novo-portal-a-caminho-da-implementacao/>>. Acesso em 25 maio 2020.

FUNARTE. Sobre a Instituição. Brasil: Associação Cultural Funarte – Fundação Nacional de Artes – Secretaria Especial da Cultura – Governo Federal. 2010. Disponível em: <<http://www.funarte.gov.br/a-funarte/>>. Acesso em 25 maio 2020.

FUNDAÇÃO CASA DE RUI BARBOSA. Histórico. Brasil: Fundação Casa Rui Barbosa. 2018b. Disponível em: <http://www.casaruibarbosa.gov.br/interna.php?ID_S=73>. Acesso em 25 maio 2020.

FUNDAÇÃO CASA DE RUI BARBOSA. Missão. Brasil: Fundação Casa Rui Barbosa. 2018c. Disponível em: <http://www.casaruibarbosa.gov.br/interna.php?ID_S=10>. Acesso em 25 maio 2020.

FUNDAÇÃO CASA DE RUI BARBOSA. Sobre a Fundação. Brasil: Fundação Casa Rui Barbosa. 2018a. Disponível em: <http://www.casaruibarbosa.gov.br/interna.php?ID_S=1>. Acesso em 25 maio 2020.

FUNDAÇÃO CULTURAL PALMARES. Apresentação. Brasil: Fundação Cultural Palmares - Governo Federal. 2016. Disponível em: <http://www.palmares.gov.br/?page_id=95>. Acesso em 25 maio 2020.

IBRAM. Estrutura. Brasil: Instituto Brasileiro de Museus – Secretaria Especial da Cultura – Governo Federal. 2018a. Disponível em: <<https://www.museus.gov.br/estrutura-do-ibram/>>. Acesso em 25 maio 2020.

IBRAM. O Ibram. Brasil: Instituto Brasileiro de Museus – Secretaria Especial da Cultura – Governo Federal. 2018b. Disponível em: <<https://www.museus.gov.br/acessoainformacao/>>. Acesso em 25 maio 2020.

IPHAN. O Iphan. Brasil: Instituto do Patrimônio Histórico e Artístico Nacional – Secretaria Especial da Cultura – Governo Federal. 2014b. Disponível em: <<http://portal.iphan.gov.br/pagina/detalhes/872>>. Acesso em 25 maio 2020.

IPHAN. Rede de Arquivos Iphan. Brasil: Instituto do Patrimônio Histórico e Artístico Nacional – Secretaria Especial da Cultura – Governo Federal. 2014a. Disponível em: <<http://portal.iphan.gov.br/pagina/detalhes/1223>>. Acesso em 25 maio 2020.
maio 2020.

PUNTONI, Pedro. Rede Memorial: cultura digital, redes colaborativas e a digitalização dos acervos memoriais no Brasil. In: GOBEL, Barbara, CHICOTE, Gloria (org.). Transiciones inciertas: archivos, conocimientos y transformación digital en América Latina. La Plata.: Universidad Nacional de la Plata. Faculdade de Humanidades y Ciencias de la Educación; Berlin: Ibero-Amerikanisches institut. 2017.

SECRETARIA ESPECIAL DA CULTURA. Competências. Brasil: Governo Federal. 2013a. Disponível em: <<http://cultura.gov.br/acesso-a-informacao/institucional/competencias/>>. Acesso em 25 maio 2020.

SECRETARIA ESPECIAL DA CULTURA. Entidades Vinculadas. Brasil: Governo Federal. 2013c. Disponível em: <<http://cultura.gov.br/secretaria/entidades-vinculadas/>>. Acesso em 25 maio 2020.

SECRETARIA ESPECIAL DA CULTURA. Histórico. Brasil: Governo Federal. 2013b. Disponível em: <<http://cultura.gov.br/acesso-a-informacao/institucional/historico/>>. Acesso em 25 maio 2020.

SECRETARIA ESPECIAL DA CULTURA. Institucional. Brasil: Governo Federal. 2013d. Disponível em: < <http://cultura.gov.br/acesso-a-informacao/institucional>>. Acesso em 25 maio 2020.

SHINTAKU, Milton. Tecnologias para Gestão da Informação. 2017. p. 65-89. Repositórios digitais: teoria e prática / organização de Fernando Vechiato ... [et al.]. – Curitiba: EDUTFPR, 2017. 271 p.: il. ; 23 cm. ISBN: 978-85-7014-197-2 (E-book). Disponível em: <<http://repositorio.utfpr.edu.br/jspui/>>. Acesso em 13 nov 2018.