

# O Corpus M.A.P.:

**Desenvolvimento de um sistema de edições  
filológicas digitais para a documentação sobre as  
Mulheres na América Portuguesa**

# Relatório Científico

**Fevereiro, 2025**

**Maria Clara Paixão de Sousa  
Universidade de São Paulo**

Apresentação.....	1
1 Balanço de metas e resultados.....	1
1.1 Balanço geral amplo, 2021-2025	1
1.2 Balanço pontual em três períodos	2
1.3 Outros resultados	4
2 Detalhamento .....	5
2.1 Prática filológica	5
2.2 Desenvolvimento computacional	9
2.3 Outras ações, produtos e resultados	18
3 Fechamento e perspectivas.....	19
Anexos .....	21

# Apresentação

Este Relatório narra os trabalhos junto ao Projeto **O Corpus M.A.P.: Desenvolvimento de um sistema de edições filológicas digitais para a documentação sobre as Mulheres na América Portuguesa** entre **agosto de 2022 e janeiro de 2025**, com fomento da Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP.

O Relatório comprehende o período de vigência regular planejado para o fomento (agosto de 2022 a julho de 2024) e o período de prorrogação concedido pela Fapesp (agosto de 2024 a janeiro de 2025). Nesse sentido, lembramos que, além do relatório parcial enviado em julho de 2023, foi enviado também um relatório intermediário, em maio de 2024 (portanto a dois meses do término inicialmente previsto), no qual solicitamos prorrogação de prazo para janeiro de 2025. No presente Relatório faremos referência a esse relatório intermediário quando pertinente, tendo em vista que na ocasião, além de reportar as atividades realizadas entre agosto de 2023 e maio de 2024, sugerimos novas metas para os meses de prorrogação. Dessa forma, embora o espírito central do presente Relatório seja o de fazer um balanço dos avanços e dificuldades dos trabalhos realizados desde o início do fomento, damos especial atenção às tarefas realizadas no período da prorrogação gentilmente concedida.

Dessa forma, em **1 Balanço geral**, relatamos resumidamente os progressos da pesquisa frente às metas estabelecidas (tanto inicialmente como em reformulações subsequentes), em **2 Detalhamento** descrevemos detidamente as atividades principais do período recente, e em **3 Fechamento e perspectivas** fazemos uma apreciação geral dos trabalhos no Projeto no período da vigência deste fomento, mostrando horizontes para seu prosseguimento futuro.

Saliente-se que o corpo do documento é integrado ainda por **Anexos**, compostos dos relatórios sintéticos dos bolsistas. Completa o Relatório, por fim, um **Apêndice Digital**, contendo entre outros, as listas completas dos manuscritos editados e as ligações para os produtos mais relevantes citados no texto, disponível em [map.prp.usp.br/Relatorio/2025.html](http://map.prp.usp.br/Relatorio/2025.html).

## 1 Balanço de metas e resultados

### 1.1 Balanço geral amplo, 2021-2025

Iniciamos este Relatório com um balanço entre as metas estabelecidas na Proposta do Projeto, em dezembro de 2021 (para início em agosto de 2022), e os resultados alcançados em janeiro de 2025. Em 2021, os resultados esperados do trabalho foram assim definidos (no quadro-resumo da seção **2. Resultados esperados** da Proposta enviada à Fapesp em dezembro de 2021, aqui reproduzido):

#### Quadro-resumo dos resultados esperados (dezembro, 2021)

1. O **Corpus M.A.P. Beta**, com **40** manuscritos editados eletronicamente;
2. Metodologias e ferramentas computacionais:
  - 2.1 Sistemas de metadados para **Catálogo** e **Corpus M.A.P.**;
  - 2.2 O **eDíctor 2.0**;
  - 2.3 Modelos treinados para leitura automática de manuscritos (**HTR**);
3. Formação e capacitação pessoal para a continuidade do Projeto.

Em janeiro de 2025, o quadro de resultados obtidos apresenta-se da seguinte forma:

## Quadro-resumo de resultados obtidos (janeiro, 2025)

1. O **Corpus M.A.P.**, com 94 manuscritos editados eletronicamente;
2. Metodologias e ferramentas computacionais:
  - 2.1 Um sistema de metadados único e integrado para Catálogo e Corpus, compondo a base do **Portal M.A.P.**; a arquitetura do Portal M.A.P. desenvolvida; a interface e o funcionamento dinâmico do Portal M.A.P. parcialmente desenvolvidos;
  - 2.2 O **eDictor 2.0** desenvolvido, mas não disponível publicamente;
  - 2.3 Modelos treinados para leitura automática de manuscritos (**HTR**) desenvolvidos e aplicados a um subcorpus, em experimento a ser publicada em prestigioso periódico internacional;
3. Formação e capacitação pessoal para a continuidade do Projeto.

Os resultados atendem às metas iniciais de forma variada, com avanços para além do esperado (em particular no trabalho de prospecção documental e edição filológica de manuscritos), um resultado obtido com atraso (o eDictor 2.0), um resultado obtido antes do esperado (os modelos de HTR, prontos já desde o primeiro ano de pesquisas), e uma reformulação de metas quanto aos produtos principais. Nesse caso, centralmente, a meta inicial de publicar o **Corpus M.A.P. beta** foi superada pelo desenvolvimento do **Portal M.A.P.**; de modo relacionado, a meta de desenvolver 'sistemas de metadados para **Catálogo** e **Corpus**' foi superada pelo desenvolvimento de metadados integrados, como fundamento do Portal.

Assim, para um balanço mais preciso, será importante retomar pontualmente a evolução das metas e resultados parciais ao longo da vigência do fomento, lembrando os trabalhos em três períodos: entre o início da vigência do fomento e o relatório parcial (agosto/2022 – julho/2023), entre o relatório parcial e o relatório intermediário (agosto/2023 – maio/2024) e este período final de prorrogação, entre o relatório intermediário e o presente (junho/2024 – janeiro/2025).

## 1.2 Balanço pontual em três períodos

### 1.2.1 Da proposta inicial ao primeiro relatório (agosto/2022 – julho/2023)

Em julho de 2023, reportamos que o primeiro ano de pesquisas havia sido dedicado ao adensamento do trabalho filológico e computacional, o que incluiu o adiantamento do desenvolvimento de alguns métodos que estavam previstos apenas para o segundo ano (os experimentos com a tecnologia HTR), conforme explicita o relatório então enviado e aprovado. Além disso, reportamos que havíamos concentrado as atividades daquele primeiro ano nos documentos de que já dispúnhamos, frutos de pesquisas anteriores ao Fomento, no Arquivo Público do Estado de São Paulo (APESP), no Instituto de Estudos Brasileiros (IEB) e no acervo digital do Arquivo Nacional da Torre do Tombo (ANTT), não tendo havido atividade significativa de prospecção de novos documentos no período. Ainda naquele relatório, previu-se para o segundo ano uma intensificação da atividade de prospecção de documentos e a finalização do desenvolvimento dos produtos principais da pesquisa:

**Metas estabelecidas em julho/2023 (relatório parcial) para agosto/2023 a julho/2024:**

- *Prospecção de documentos e ampliação da base documental;*
- *Finalização e lançamento do **Catálogo 2.0**;*
- *Finalização e lançamento do **eDictor 2.0**;*
- *Desenvolvimento do **Catálogo 3.0**;*
- *Finalização e lançamento do **Corpus M.A.P.** com funcionalidades completas, integrado ao Catálogo 3.0.* (texto do Relatório Parcial de julho/2023)

## 1.2.2 Entre o primeiro e o segundo relatório (agosto/2023 – maio/2024)

No segundo relatório, reportamos os progressos das estabelecidas para o segundo ano, a dois meses de seu término (em maio de 2024). Como relatado então, entre agosto de 2023 e maio de 2024 os bolsistas e pesquisadores não-bolsistas dedicaram-se a ampliar a base documental do Projeto e finalizar os produtos e métodos previstos. Observamos na ocasião que o sucesso dessas atividades havia sido variado: de um lado, tínhamos obtido progressos significativos na prospecção documental e no desenvolvimento das metodologias de catalogação e edição; de outro lado, dois produtos não haviam sido finalizados conforme o Cronograma, e não poderiam ser lançados até julho de 2024 (prazo original do término do fomento). Especificamente, naquele relatório pontuamos os sucessos e desafios do período da seguinte forma:

**Metas estabelecidas em julho/2023 alcançadas até maio/2024 (relatório intermediário):**

- **Base documental ampliada** dos 40 documentos previstos no projeto inicial para **100** documentos. Desses, **31** são novos manuscritos editados ou em processo de edição e **45** são novos documentos prospectados em Arquivos (já sistematizados, a serem editados e incluídos em versões futuras), além das **24** edições filológicas publicadas como 'Coleções Preliminares' em 2023;
- **A versão 2.0 do Catálogo** (ou seja, com o novo desenvolvimento em Python e visualização em formato de banco de dados) em fase de testes de usabilidade a partir da importação da versão 1.0 (ou seja, a versão ainda com visualização XML apresentada no relatório anterior). (texto do Relatório Intermediário de Maio/2024)

**Metas estabelecidas em julho/2023 não alcançadas até maio/2024 (relatório intermediário):**

- O **eDictor 2.0** não está pronto. O novo cronograma do trabalho de tese da pesquisadora não-bolsista Aline Costa indica o término em agosto de 2024 (...);
- O '**Catálogo 3.0**' não está publicado. Nesse caso, tratava-se de uma tarefa dos bolsistas da equipe, diretamente impactada pela mudança conceitual na nossa perspectiva sobre a arquitetura final do Projeto. (texto do Relatório Intermediário de Maio/2024)

Conforme observamos naquele relatório intermediário, a diferença entre o cronograma previsto e as tarefas efetivamente terminadas até então explicava-se em parte pela ação de contingências alheias à nossa vontade (caso do eDictor 2.0) e em parte pela reavaliação da nossa metodologia de trabalho (caso da eliminação da planejada versão 3.0 do Catálogo). Na ocasião, detalhamos as perspectivas diante dessas contingências e dessa mudança conceitual. Salientamos então que até o período compreendido no relatório intermediário, a equipe havia se dedicado ao aprimoramento da estrutura formal do **Catálogo M.A.P** e à ampliação da base documental do **Corpus M.A.P Versão Beta**; entretanto, como também relatamos, o amadurecimento das pesquisas nos levara a conceber uma nova forma de funcionamento para os produtos finais do Projeto. Mais que publicar Catálogo e Corpus como objetos separados (embora interligados), almejávamos organizar um ambiente digital constituído organicamente da união de 'Catálogo' (isso é, das informações sobre os documentos) e 'Corpus' (isso é, das edições dos documentos). Esse espaço, a que denominamos então o **Portal M.A.P.**, traria vantagens tanto para o futuro das pesquisas do grupo como para os usuários externos, permitindo uma experiência de leituras e pesquisas mais ágil e (ao mesmo tempo) mais controlada do que a anteriormente prevista. Diante dessas mudanças, apresentamos aquele relatório antes do prazo final, sugerindo que o encerramento do fomento fosse adiado por seis meses. Tal sugestão foi gentilmente concedida pela Fapesp. Para atender a essa reformulação, sugerimos na ocasião as seguintes metas para os seis meses de prorrogação do fomento:

**Metas estabelecidas em maio/2024 para junho/2024 até janeiro/2025 (período de prorrogação):**

- **Desenvolvimento do Portal M.A.P.**, espaço que unirá os documentos filologicamente editados e o banco de dados em uma mesma base e interface. O desenvolvimento pressupõe a **adaptação das**

edições digitais ao novo sistema, e o desenho de uma interface orgânica, a partir das duas interfaces separadas existentes hoje;

- **Preparo do Manual M.A.P.**, uma documentação detalhada e abrangente, que integrará desde as decisões pertinentes ao trabalho de edição até a organização digital da base documental, e será parte integrante do Portal – destinando-se tanto a futuros pesquisadores do grupo como ao usuário final do sistema. (texto do Relatório Parcial de Maio, 2024)

### 1.2.3 Do segundo relatório até o presente (junho/2024 – janeiro/2025)

Presentemente, o balanço pontual das duas metas principais estabelecidas para o período da prorrogação é como segue:

#### Metas estabelecidas em maio/2024 alcançadas até janeiro/2025:

- **Publicação de uma versão inicial do Portal M.A.P.**, espaço que une os documentos filologicamente editados e o banco de dados em uma mesma base. A **adaptação dos metadados e das edições digitais ao novo sistema** foi desenvolvida, mas o **desenho de uma interface orgânica** ainda não está satisfatório, por razões detalhadas à frente neste Relatório;
- **Publicação do Manual M.A.P.**, documentação que integra desde as decisões pertinentes ao trabalho de edição até a organização digital da base documental, e é parte integrante do Portal, destinada a futuros pesquisadores do grupo e ao usuário do sistema.

Por fim, mas muito importante, ressaltamos que a equipe intensificou sua dedicação aos trabalhos de prospecção documental e edição filológica nesse período final. Assim, a base documental do Projeto está novamente ampliada com relação a 2024 (não tendo sido essa uma meta específica para a prorrogação, e sim uma tarefa constante em nossas pesquisas):

- **Base documental ampliada:** A base do Projeto saltou para **718** documentos em diferentes estágios de processamento: **94** manuscritos editados filologicamente publicados, **82** documentos plenamente descritos constantes do Catálogo mas ainda sem edição filológica publicada, e **542** documentos prospectados e ainda não plenamente catalogados. Salienta-se o salto da produção no período da prorrogação (**63** novas edições publicadas e **450** novos documentos prospectados), e o volume total de texto editado: em conjunto, os manuscritos com edições publicadas ao longo da vigência somam **866 fólios** e cerca de **144.000** palavras.

Na **seção 2** à frente descrevemos em detalhe os resultados no que toca à ampliação da base documental e também ao trabalho de desenvolvimento computacional mencionado acima.

### 1.3 Outros resultados

Dois outros resultados importantes tiveram progressos no período estendido do fomento Fapesp:

- **Aceite da submissão de artigo sobre os modelos de HTR e tratativas para convênio com a Universidade de Innsbruck.** Citado como pendente de avaliação no relatório anterior, o artigo *Digital technologies for Early Modern Portuguese manuscripts: an experiment with two Handwritten Text Recognition software applications* foi aceito para publicação no periódico **Digital Scholarship in the Humanities**, da Universidade de Oxford. Estamos também em tratativas para nos tornarmos membros da **ReadCoop**, sediada na **Universidade de Innsbruck**, Áustria, responsável pelo software **Transkribus** (cf. 2.3);
- **O eDictor 2.0 ficou pronto** em dezembro de 2024 (cf. 2.3.2).

Listamos, por fim, a produção científica coletiva diretamente ligada ao Projeto:

1. Cardenete, Beatriz de Freitas, Motta, Elisa Hardt Leitão; Monte, Vanessa Martins do. O estudo diplomático de manuscritos de e sobre mulheres na América Portuguesa: cartas, respostas e informações. *Filologia E Linguística Portuguesa*, 26(1), 9-27, 2025. <https://doi.org/10.11606/issn.2176-9419.v26i1p9-27>
2. Costa, Aline Silva; Namiuti, Cristiane; Paixão de Sousa, Maria Clara; Costa, Bruno Silvério; Santos, Jorge Viana. An annotation proposal based on TEI Schema for Portuguese Corpora editions: A solution for e-Dictor XML annotation problem. *Proceedings of the Second Workshop on Digital Humanities and Natural Language Processing (2nd DHandNLP 2022)*, 2022. <https://ceur-ws.org/Vol-3128/paper7.pdf>
3. Motta, Elisa Hardt Leitão; Guets, Raquel de Paula; Vita, Mariana Rodrigues de; Paixão de Sousa, Maria Clara; Monte, Vanessa Martins do. Projeto Mulheres na América Portuguesa: um processo de edição filológica eletrônica. In C. Altman; J. Lourenço (Eds.), *Feminismo em historiografia linguística: Américas*. Volume I, pp. 261-284. Campinas: Pontes Editores, 2023.
4. Motta, Elisa Hardt Leitão; Monte, Vanessa Martins do; Paixão de Sousa, Maria Clara. As mulheres nos requerimentos do Fundo da Secretaria de Governo da Capitania de São Paulo: 1800-1822. In F. Gonçalves (Org.), *História e Língua: Interfaces. Publicações do Centro Interdisciplinar de História, Culturas e Sociedades (CIdeHUS)*, Univ. de Évora. Submetido em 2023; no prelo.
5. Paixão de Sousa, Maria Clara; Monte, Vanessa. Mulheres na América Portuguesa: primeiros ecos de um passado silenciado. In K. Carula; L.G. Magalhães (Orgs.), *Escritas femininas*. Niterói, RJ: Editora da Universidade Federal Fluminense. Submetido em 2023; no prelo.
6. Paixão de Sousa, Maria Clara; Monte, Vanessa; Namiuti, Cristiane; Santos, Jorge; Xavier, Lúcia; Magalhães, Lívia; Costa, Bruno; Costa, Aline; Sturzeneker, Mariana; Crespo, Maria Clara; Rocha, Maria Lina; Zacchi, Natalia. Digital Technologies for Early Modern Portuguese Manuscripts: An Experiment with Two 'Handwritten Text Recognition' Software Applications. *Preprints* 2024, 2024120864. <https://www.preprints.org/manuscript/202412.0864/v1>
7. Paixão de Sousa, Maria Clara; Monte, Vanessa; Namiuti, Cristiane; Santos, Jorge; Xavier, Lúcia; Magalhães, Lívia; Costa, Bruno; Costa, Aline; Sturzeneker, Mariana; Crespo, Maria Clara; Rocha, Maria Lina; Zacchi, Natalia. Digital technologies for Early Modern Portuguese manuscripts: An experiment with two Handwritten Text Recognition software applications. *Digital Scholarship in the Humanities*. Submetido em 2024; aceito em janeiro de 2025 (DSH-2024-0170).

## 2 Detalhamento

### 2.1 Prática filológica

#### 2.1.1 Ampliação da base documental

A base documental do Projeto, prevista para conter 40 documentos filologicamente editados e um número aberto de documentos a serem prospectados, conta hoje com **718** documentos em diferentes estágios de processamento: **94** manuscritos plenamente catalogados e com edição filológica publicada no corpus, **82** manuscritos plenamente catalogados mas sem edição filológica publicada no corpus, e **542** manuscritos prospectados em arquivos (480 em arquivos físicos e 62 em acervos digitais) mas ainda não plenamente catalogados nem editados.

Essa ampliação da base se deve à ação incansável da equipe filológica do Projeto, em particular no período final e na prorrogação da vigência do fomento. A base documental começa a crescer já no segundo ano de trabalhos (do final de 2023 para o início de 2024), quando passamos de 24 para 31 edições filológicas publicadas e tendo sido incluídos ainda mais 45 novos documentos prospectados em Arquivos. O maior progresso aconteceu no último semestre, já no período estendido da vigência: nessa etapa, foram publicadas mais 63 novas edições e prospectados

mais 497 documentos ainda não plenamente catalogados. A **Tabela 1** resume este cenário, organizando os manuscritos segundo estágios de processamento e arquivos de origem.

**Tabela 1:** Número de manuscritos prospectados, catalogados e publicados (por arquivos de origem)

Arquivo de origem	prospectados ainda não catalogados	catalogados sem edição publicada	catalogados com edição publicada	em todos os estágios de preparação
Arquivo Público do Estado de São Paulo (APESP)	2	37	31	<b>70</b>
Arquivo Nacional da Torre do Tombo (ANTT)	62	27	27	<b>116</b>
Biblioteca Nacional do Rio de Janeiro (BNRJ)	27	2	26	<b>55</b>
Instituto de Estudos Brasileiros (IEB)	1	0	10	<b>11</b>
Arquivo Eclesiástico da Arq. de Mariana (AEAM)	450	0	0	<b>450</b>
Outros arquivos (CMJ, APM, ACMSP)	0	16	0	<b>16</b>
<b>Em todos os arquivos:</b>	<b>542</b>	<b>82</b>	<b>94</b>	<b>718</b>

Nas subseções a seguir discutimos esses progressos: em **2.1.2** detalhamos os trabalhos de prospecção em cada Arquivo, descrevendo o perfil dos documentos estudados, e em **2.1.3** tecemos uma apreciação geral dos trabalhos de edição filológica e seu impacto no Projeto.

Importa pontuar que, em contraste ao que foi feito para os Relatórios Científicos anteriores, na presente ocasião não listaremos no corpo do Relatório os documentos prospectados e editados, por uma questão de espaço. As listas da base documental, com referências completas, estão disponíveis no **Apêndice Digital** ([map.prp.usp.br/Relatorio/2025.html](http://map.prp.usp.br/Relatorio/2025.html)); mais especificamente,

- Lista de documentos prospectados: [map.prp.usp.br/Relatorio/Prospectados.html](http://map.prp.usp.br/Relatorio/Prospectados.html)
- Lista de documentos editados: [map.prp.usp.br/Relatorio/Editados.html](http://map.prp.usp.br/Relatorio/Editados.html)

## 2.1.2 Trabalho de prospecção e estudo documental por Arquivo

**Arquivo Eclesiástico da Arquidiocese de Mariana (AEAM).** Foi especialmente frutífera para a ampliação da base documental do Projeto a pesquisa realizada pelas bolsistas Andréa Natanael da Silva e Nicóllia de Lima Garcia junto ao Arquivo Eclesiástico da Arquidiocese de Mariana (AEAM), em Mariana-MG, em janeiro de 2025. Na visita, foi possível avaliar todo o acervo documental da instituição e avaliar quais fundos seriam de interesse para o Projeto. Foram prospectaram **450** documentos, a saber: **149** testamentos realizados entre os anos de 1720 a 1806; **4** cartas realizadas entre os anos de 1747 e 1781; **3** breves realizados entre os anos de 1795 a 1804; **86** termos do Livro de Devassas; **58** termos do Livro de termos de Visita à Sabará com sua Comarca (Devassas); **104** documentos provenientes do Juízo Eclesiástico entre os anos de 1718 e 1808; **9** documentos provenientes do Livro de Devassas da Freguesia de Camargo no ano de 1745; **16** documentos provenientes do livro de Devassas de 1721 a 1735; **17** documentos provenientes do Livro do Juízo Eclesiástico de 1748 a 1765; e **4** documentos provenientes do Livro do Juízo Eclesiástico de 1765 a 1784. O trabalho das bolsistas foi exemplar, e a rica documentação resultante terá destaque nas versões futuras do Catálogo e do Corpus.

**Arquivo Público do Estado de São Paulo (APESP).** A pesquisa junto ao APESP tem sido frequente desde o início do Projeto, tanto pela riqueza de seu acervo como pelo fato prático de sua proximidade geográfica. Foram prospectados ao todo **70** documentos relevantes, no **Fundo da Secretaria de Governo da Capitania de São Paulo (1721 - 1823)** - BR SPAPESP SEGOVC, “composto por documentos produzidos pelas autoridades do reino, da colônia e da capitania, além de

particulares solicitando perdão, auxílio e mercês"<sup>1</sup>, sendo os documentos de particulares especialmente ricos para o Projeto. Esta área de pesquisa é conduzida pela bolsista Elisa Hardt Leitão Motta, cuja dedicação e atuação autônoma junto ao APESP é notável. Ainda como fruto do trabalho da bolsista, **68** dos 70 documentos estão plenamente catalogados, **31** dos quais já com edições publicadas.

**Arquivo Nacional da Torre do Tombo (ANTT).** A pesquisa junto ao acervo digital do ANTT foi também uma atividade regular do Projeto desde o início, em particular no que toca ao Fundo do Tribunal do Santo Ofício – ANTT TSO<sup>2</sup>. Nos últimos anos, essa prospecção digital esteve sobretudo a cargo da bolsista Andréa Cristina Natanael, em particular quanto à seção Inquisição de Lisboa e, nela, os documentos referentes às Visitações ao Brasil (livros de registro, confissões, denúncias e processos). A bolsista Nicólli de Lima Garcia também realizou pesquisas no acervo digital, prospectando documentos de tipologias diversas (processos, outros sumários processuais e denúncias), provenientes de períodos e regiões distintas com relação à pesquisa de Andréa Natanael (documentos da segunda metade do século XVIII, e documentos originários das regiões do Pará e Maranhão). Temos ao todo **116** documentos como resultado dessas pesquisas digitais, dos quais **27** estão catalogados e editados, e **62** são achados recentes a serem processados futuramente. Neste último período, fizemos, pela primeira vez, pesquisas junto ao arquivo físico, em Lisboa, em julho de 2024. Na ocasião, as coordenadoras do Projeto puderam examinar detidamente outros fundos e seções de interesse, sobretudo no que toca materiais não digitalizados, ou com digitalização parcial ou precária. Destacam-se aí diversos códices da série **Embarques dos condenados a degredo para o Ultramar** (PT/TT/JD/010) e o **Repertório das Mulheres** (PT/TT/TSO-IL/010/0028), material de suma importância a que voltamos na [seção 3](#).

**Biblioteca Nacional do Rio de Janeiro (BNRJ).** A BNRJ foi destino de duas pesquisas in loco durante a vigência do Projeto. Em agosto de 2023, a bolsista Elisa Motta realizou uma primeira visita à Seção de Manuscritos, selecionando 27 novos conjuntos documentais para o Projeto (em especial, cartas particulares). Em janeiro de 2025, as coordenadoras do Projeto voltaram à Seção com o intuito específico de descobrir documentos relativos à atuação do Tribunal do Santo Ofício excepcionalmente incluídos na guarda da BN. É o caso, principalmente, de nove códices listados no Catálogo da BNRJ como *"Livros da Inquisição de Goa"* (Manuscritos – 25,01,001 a Manuscritos – 25,01,009). Embora constem como 'digitalizados' na base de dados da BNRJ, os códices são imensos e a digitalização remete a alguns poucos documentos de cada volume. Foi extremamente importante fazer a pesquisa minuciosa desses livros in loco, pois eles encerram grande potencial para o Projeto. Os códices são compostos por comunicações administrativas entre o Tribunal em Portugal e sua seção em Goa, oferecendo acesso inédito, para nós, ao universo de procedimentos internos do TSO. No futuro, pretendemos realizar mais pesquisas nesse conjunto da BN, em complementação às pesquisas digitais e no fundo central em Portugal.

**Instituto de Estudos Brasileiros (IEB).** Ao longo do Projeto houve também pesquisas regulares no IEB. Em um primeiro momento, destacou-se o trabalho da pesquisadora não-bolsista Beatriz de Freitas Cardenete, que contribuiu com a edição de 9 cartas pessoais provenientes de suas pesquisas nesse Acervo, como parte de seu trabalho de doutorado. Além disso, a bolsista Nicólli Garcia e a profa. Vanessa do Monte realizaram uma extensa pesquisa na base de dados do IEB, abrangendo a **Coleção Alberto Lamego**, cujas data-limite coincidem, em grande parte, com as

<sup>1</sup> Cf. Descrição arquivística em <http://icaatom.arquivoestado.sp.gov.br/ica-atom/index.php/secretaria-de-governo-da-capitania-de-sao-paulo>

<sup>2</sup> Cf. Descrição arquivística em <https://digitarq.arquivos.pt/details?id=2299703>

do Projeto M.A.P. Nesse levantamento, conseguiram encontrar apenas um único documento que poderia ser incluído em nosso corpus, mas que ainda será preciso consultar e editar in loco.

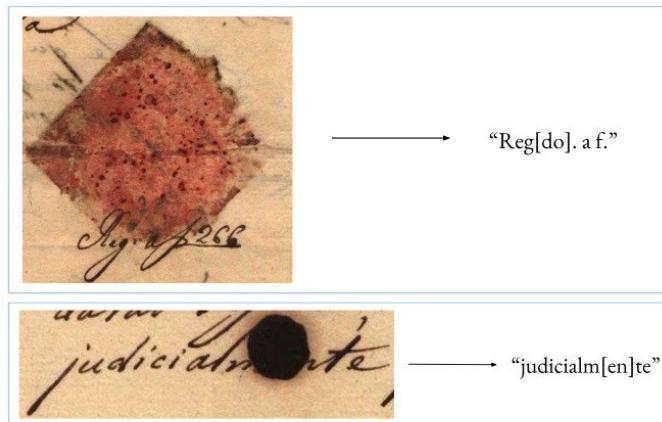
### 2.1.3 Trabalho de edição, reflexão e análise filológica

Mais que um aumento quantitativo, a ampliação da base documental do Corpus trouxe avanços na qualidade do trabalho – em particular, por nos permitir ampliar a variedade tipológica de documentos estudados. Essa ampliação foi fruto os trabalhos das bolsistas diretamente envolvidas com as atividades filológicas – Andréa Natanael da Silva, Elisa Hardt Leitão Motta e Nicóllia de Lima Garcia, responsáveis por resultados importantes do Projeto nas áreas de pesquisa filológica, prospecção e edição de documentos.

A começar pela questão da ampliação de escopo tipológico, destacam-se o trabalho de Elisa Motta sobre uma série de requerimentos de mulheres dirigidos ao Ministério do Império e o trabalho de Nicóllia Garcia sobre o sumário-crime contra Antônia Maria de Almeida – documentos que garantirão ao corpus uma desejável variedade tipológica, enriquecendo o estudo diplomático dos manuscritos. O caso do sumário-crime é exemplar: tendo sido elaborado entre a América Portuguesa e o reino, o manuscrito permite análises aprofundadas da composição e autoria de suas partes; do modo como se conduziam as investigações das denúncias do TSO, já perto de seu fim; e do tempo que levava o envio e o retorno dos papéis que registravam os supostos erros de fé. Por sua vez, o conjunto de requerimentos prospectados por Elisa Motta na BNRJ traz uma oportunidade de contraste tipológico com os requerimentos do APESP. Apesar de serem tratados arquivisticamente como “requerimentos”, a pesquisa de doutorado de Elisa vem mostrando tratarem-se de documentos peticionários, muitas vezes compostos por outras tipologias. Assim, o incremento documental, além de quantitativo, é também qualitativo, permitindo estudos diplomáticos muito ricos.

Importa sublinhar como, a partir de seus trabalhos individualizados, as bolsistas desenvolveram alguns objetivos próprios que valem ser incentivados, tanto por demonstrar amadurecimento acadêmico e iniciativa de investigação como por representarem ganhos potenciais para o Projeto coletivo. Nesse sentido, Elisa Motta tem se destacado por seu interesse no desenvolvimento da metodologia para a edição filológica modernizada, parte central da proposta do Projeto; Andréa Natanael e Nicóllia Garcia vêm demonstrando especial interesse pela questão das abreviaturas características dos documentos inquisitoriais, e pretendem realizar estudos visando a classificação e investigação de sua presença em dicionários especializados.

O descriptor **manchado** é utilizado para sinalizar palavras que se apagaram ou estão parcialmente visíveis por mancha, seja de tinta respingada ou tinta vazada do verso do fólio, por tinta de carimbos ou cola de selos:



Fragmentos do Sumário Crime de Antônia Maria de Almeida  
(ANTT - PT/TT/TSO-IL/028/09738)

**Figura 1.** Trecho do Manual de Edição ilustrando parte da descrição diplomática e paleográfica da equipe

Além de seu potencial futuro, essas três dimensões trazidas pela motivação particular das bolsistas (estudo diplomático, investigação paleográfica, proposta metodológica para edição em camadas) já se refletem em benefícios imediatos para o Projeto. Em particular, tiveram imenso impacto na qualidade filológica do **Manual de Edições** recentemente publicado (cf. 2.2.2). O cuidado e a minúcia das bolsistas ao descrever e discutir suas opções editoriais na preparação dessa documentação, que representa um resultado central desse período, foi notável.

## 2.2 Desenvolvimento computacional

### 2.2.1 Concepções do M.A.P.: de um Catálogo a um Corpus e a um Portal

#### 2.2.1.1 Breve histórico

Damos início ao detalhamento dos trabalhos de desenvolvimento computacional nesses últimos meses com um breve histórico dos horizontes tecnológicos do M.A.P., retomando a trajetória na qual partimos do objetivo de construir um Catálogo, expandimos esse objetivo para a construção de um Corpus ligado ao Catálogo e, por fim, idealizamos construir um espaço digital integrado de documentos editados e descritos, a que chamamos o Portal M.A.P. Idealizado em 2017, o M.A.P. inicialmente almejava apenas compor um Catálogo, com recursos tecnológicos limitados, mas adequados para nossos objetivos originais<sup>3</sup>:

Neste projeto, iremos buscar e catalogar documentos escritos por mulheres no contexto da formação da América Portuguesa, de modo a torná-los visíveis para pesquisas futuras nos campos da história do Brasil e da Filologia Portuguesa, seja em nossas próprias investigações, seja nas de outros pesquisadores desses dois campos. Como produto principal da pesquisa, ofereceremos à público, em ambiente digital com acesso aberto, **um catálogo sistematizado** com informações arquivísticas e temáticas sobre cada documento encontrado, e **um índice onomástico** das mulheres escreventes e das mulheres com discurso relatado nos documentos. (Trecho do projeto original, de 2017)

O passar dos anos mostrou que a complexidade da documentação reunida nos obrigava a adensar o tratamento tecnológico da base documental, e a riqueza da documentação nos impelia a produzir um Corpus Eletrônico filologicamente editado – o que começamos já em 2019, lançando uma primeira edição (a que chamamos ‘piloto’). O início do fomento Fapesp em 2022 foi o ponto de inflexão crucial nesse sentido; ali efetivou-se o passo de construir um Corpus associado ao Catálogo, como se vê no seguinte trecho da Proposta inicial à Fapesp:

A meta do projeto aqui proposto é a de publicar o **Corpus M.A.P. Beta** – versão inicial de um corpus inédito formado por manuscritos de mulheres e sobre mulheres da América Portuguesa, editados filologicamente com tecnologias digitais de última geração, e com interesse para pesquisas nos âmbitos da filologia, da história da língua, da história social, da história da escrita e da leitura, e da história das mulheres no Brasil. (...) O Corpus M.A.P. Beta será composto por 40 documentos manuscritos editados eletronicamente (texto da Proposta enviada em dezembro de 2021)

Os primeiros dois anos de formação do Corpus trouxeram uma nova ampliação de escopo: concebemos, em 2024, uma nova forma de funcionamento para o M.A.P., em um ambiente digital constituído organicamente da união de *Catálogo* (isso é, das informações sobre os documentos) e *Corpus* (isso é, das edições dos documentos):

Em outras palavras, não haverá *um Catálogo M.A.P. integrado a um Corpus M.A.P.*: haverá o **Portal M.A.P.**, um espaço povoado com documentos dotados de metadados ricos, no qual tudo poderá ser pesquisado sempre em contexto. (...) Consideramos essa opção conceitualmente adequada, tendo em vista que o mote principal de nossa pesquisa é oferecer aos usuários um conjunto documental reunido graças a uma curadoria delicada e cuidadosa, com material ricamente contextualizado e explicado em sua importância histórica, social e linguística. (texto do relatório de 2024).

Assim, a ideia do Portal surge como amadurecimento das nossas experiências de curadoria documental e construção de corpora, mas tem também sua raiz na nossa percepção dos limites dos recursos tecnológicos que havíamos desenvolvido até então. Já por ocasião do relatório de

<sup>3</sup> Paixão de Sousa, Maria Clara; Monte, Vanessa Martins do. ‘Agora andam me jurando a pele’: Escritos de mulheres e escritos sobre mulheres na América Portuguesa. Projeto de pesquisa, Programa Unificado de Bolsas (PUB) / Universidade de São Paulo. São Paulo; 2017. [http://map.prp.usp.br/MAP\\_Projeto\\_2017.html](http://map.prp.usp.br/MAP_Projeto_2017.html)

2024 apresentamos uma reflexão autocrítica nesse sentido, afirmando que manter o Catálogo e o Corpus como objetos separados não era uma boa prática de curadoria nem uma boa prática de organização digital da informação:

Isso pode ser observado na seguinte situação hipotética, possível na arquitetura original: um usuário final poderia ler e pesquisar um dos documentos do Corpus sem ter acesso à rica informação catalográfica que o contextualiza e explica. É de fato o que acontece, presentemente, com os documentos das *Coleções Preliminares* (lançados em 2023): o espaço das coleções (nossa atual 'Corpus') apresenta os documentos individuais independentes do Catálogo, de modo que todo seu contexto se perde na leitura e trabalho do usuário. (...) Há portanto um duplo eixo de princípios aos quais pretendemos agora nos adequar: princípios de curadoria e contextualização emanados do fundo filológico de nosso trabalho, e princípios de ética e boas práticas de tratamento da informação, ligados ao lado computacional de nossas pesquisas. (texto do relatório de 2024).

O cerne da concepção do Portal, de um ponto de vista técnico, é a integração entre as informações relativas à curadoria e descrição documental e as edições filológicas. Os meses de trabalho transcorridos desde essa concepção nos permitiram chegar a um resultado inicial interessante, mas que ainda não consideramos definitivo.

### 2.2.1.2 Aspecto atual do Portal

No que configura sua característica mais fundamental, o Portal M.A.P. já existe em versão preliminar: os 94 manuscritos do M.A.P. com edição filológica digital estão organicamente ligados à sua descrição catalográfica por meio de metadados integrados, e essas duas esferas de informação, antes separadas, já podem ser acessadas em um só ambiente digital. Uma mesma página de entrada leva a todos os 176 documentos catalogados e suas eventuais edições, não persistindo a situação anterior na qual as informações catalográficas e as edições filológicas compunham sistemas distintos, com entradas distintas e arquivos sem diálogo.



Figura 2. Página de entrada atual do Portal ([map.prp.usp.br/MAP.html](http://map.prp.usp.br/MAP.html))

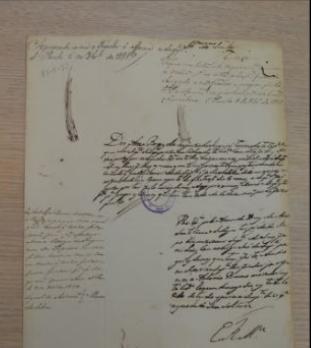
A **Figura 2** mostra o aspecto atual dessa entrada.

Os conteúdos podem ser acessados ao se clicar cada instância nomeada na nuvem de palavras, ao serem encontrados na caixa de buscas, ou por meio do menu lateral +, que ativa uma lista onomástica.

Ana Borges

Ana Borges, da freguesia de Juqueri, termo desta cidade

1819



Ana Borges requereu pela própria soltura após passar oito dias na cadeia da cidade junto de suas duas filhas. Segundo a suplicante, padeciam "as maiores misérias por não ter o que gastar" e seu marido também não as podia auxiliar por encontrar-se "alcançado em anos no fundo de uma cama há perto de quatro para cinco anos". Presa por "vários danos a seus próximos", Ana ofereceu um fiador que se responsabilizasse a pagar "tanto em moeda como em trastes que tenha furtado e possa furtar" o que a suplicante não podia, no momento, restituir. Antônio Bueno, morador na mesma cidade, "homem de negócio e estabelecido de tudo" foi apontado como o fiador de Ana e suas filhas. Por fim, a soltura da suplicante foi ordenada em um despacho

Edição diplomática  
Edição modernizada  
Texto sem marcação (txt)  
Hiperedição (xml cru)

Edições filológicas feitas por  
Elisa Hardt Leitão Motta

Figura 3. Ficha para uma catalogada ([map.prp.usp.br/0079](http://map.prp.usp.br/0079))

Mais que uma nova 'cara' para o conteúdo, essa interface corresponde a uma nova estrutura de organização de dados. De fato, o aspecto mais importante do novo formato é que a cada instância de catalogação corresponde uma estrutura de metadados (em XML) com todas as categorias descritivas do M.A.P. – e principalmente: nos documentos editados filologicamente, o documento XML que é a base da edição contém, também, esse cabeçalho de metadados com as informações catalográficas.

Assim, consideramos que a interface atual é condizente com o conceito original do Portal, uma vez que ela logra integrar organicamente essas duas ordens de informação. De outro lado, reputamos essa interface como provisória, uma vez que ela não faz essa integração dinamicamente a partir do banco de dados, mas sim ainda estaticamente, com documentos XML e HTML previamente gerados. Nas seções a seguir isso é explicado com mais detalhe.

### 2.2.1.3 Caminhos do Corpus ao Portal

Para melhor detalhar a seguir os caminhos seguidos na direção da implementação do Portal, os resultados iniciais obtidos e os desafios que encontramos, importará retomar a previsão inicial do trabalho a ser feito. Lembramos que junto com a concepção do Portal, o relatório anterior indicou as principais ações para o início de seu desenvolvimento:

- A. Adaptação dos metadados do Corpus.** Em consonância com o cerne da ideia do Portal, previmos como a primeira tarefa principal a adaptação dos metadados do Corpus aos metadados do Catálogo, criando assim uma estrutura única que unisse as duas instâncias;
- B. Adaptação da interface do Catálogo.** A segunda tarefa principal prevista era a concepção de recursos de visualização unificados dessas informações para o usuário final;
- C. Preparo de documentação: o Manual M.A.P.** Por fim, previmos que a complexidade do desenvolvimento desse novo objeto tornaria imprescindível a disponibilização de documentação detalhada, o que sugerimos fazer na forma de um manual.

A adaptação dos metadados e o preparo da documentação chegaram a resultados muito satisfatórios, enquanto o trabalho com a interface enfrentou importantes obstáculos técnicos. A seguir, detalhamos os progressos e desafios das três tarefas ao longo da prorrogação.

A entrada única leva à ficha M.A.P. de cada instância (como mostra a **Figura 3**).

A ficha contém todas as informações catalográficas e as ligações para as edições filológicas, quando disponíveis.

## 2.2.2 Detalhamento dos principais resultados e desafios

### 2.2.2.1 Adaptação dos metadados

Na proposta de reorganização de Corpus e Catálogo na forma de um Portal, afirmamos que

... a riqueza dos metadados do Catálogo está em forte contraste com a simplicidade dos metadados das edições; a ideia será integrarmos os primeiros aos últimos. Essa integração será um fator fundante da integração orgânica entre Catálogo e Corpus – de fato, tornando Catálogo e Corpus um único objeto lógico. (texto do relatório de 2024)

De fato, o contraste entre a riqueza descritiva do Catálogo e a relativa falta de informações nos arquivos do Corpus era uma característica pouco feliz de nossos resultados iniciais; por outro lado, notamos que estabelecer dois sistemas de metadados diferentes (como incialmente previsto) seria arriscado e pouco elegante. Começamos então o desenvolvimento do Portal M.A.P. por esta tarefa de integrar os metadados do Catálogo aos metadados das edições do Corpus – enfrentando de início muitas dificuldades, arquitetando versões experimentais trabalhosas, pouco flexíveis e inadequadas de modo geral.

A solução para a integração veio alguns meses depois da concepção do Portal M.A.P., quando decidi abordar a tarefa pelo avesso: tomar como ponto de partida a estrutura de metadados do Corpus gerada pelo eDictor e adaptar a ela a estrutura nativa do M.A.P.

Essa solução simples tem como base o espírito da estrutura de metadados do eDictor, concebida para ser infinitamente adaptável (cf. Trippel e Paixão de Sousa, 2006<sup>4</sup>). Pesou aqui também uma contingência prática: como o eDictor é a base de todas as edições filológicas feitas no M.A.P., para todo manuscrito editado temos sempre, de qualquer maneira, um documento XML com a estrutura de metadados automaticamente gerada pelo eDictor. No início, trabalhávamos com esses dois objetos paralelos, que não dialogavam: o XML do Catálogo e os XMLs dos documentos do eDictor. Agora, o XML do Catálogo integrou-se aos metadados eDictor dos documentos editados. Para isso, desenvolvemos uma adaptação na qual todas as categorias incluídas na descrição catalográfica do M.A.P. se tornam compatíveis com a estrutura do eDictor, sendo o mapeamento entre as duas estruturas facilitado.

```
<?xml version='1.0' encoding='utf-8'?>
<document>
<head id="FL">
<!--MAP metadata-->
  <metadata generation="MAP">...
  </metadata>

<!--edictor metadata-->
  <metadata generation="edictor_internal">...
  </metadata>
</head>
<body>...
</body>
</document>
```

Figura 4. Vista esquemática de um XML inteiro: cabeçalho ('head') e corpo do texto ('body') – Documento 0003, Denúncia contra Francisca Luís.

O XML do eDictor gera automaticamente, no cabeçalho **head**, uma categoria **metadata**; a **Figura 4** mostra um documento XML inteiro com as categorias básicas em colapso:

```
<?xml ... >
<document>
  <head id="(...)">
    <metadata generation="edictor_internal"> ...
    </metadata>
  </head>
  <body>
    <text> ...
    </text>
  </body>
</document>
```

<sup>4</sup> Trippel, Thorsten, and Paixão de Sousa, Maria Clara. Metadata and XML standards at work: a corpus repository of Historical Portuguese texts. V International Conference on Language Resources and Evaluation (LREC 2006). 2006. [http://www.lrec-conf.org/proceedings/lrec2006/pdf/378\\_pdf.pdf](http://www.lrec-conf.org/proceedings/lrec2006/pdf/378_pdf.pdf).

```
<?xml-stylesheet href="" type="text/xsl"?>
<document>
  <head id="GP">
    <metadata generation="edictor_internal">
      <meta>
        <n>Document Name</n>
        <v>GP</v>
      </meta>
      <meta>
        <n>XML generated by</n>
        <v>E-Dictor-v1.0.b010</v>
      </meta>
      <meta>
        <n>Last Saved Date</n>
        <v>31.08.2023</v>
      </meta>
      <meta>
        <n>Word Count</n>
        <v>2670</v>
      </meta>
    </metadata>
  </head>
</document>
```

Figura 5. Metadados automáticos do eDictor

No caso do M.A.P., aproveitamos o recurso das ‘gerações’ de metadados do eDictor para inserir nos documentos uma categoria **metadata** com o atributo **generation="MAP"**, que contém 80 camadas **meta** com as informações catalográficas do M.A.P. (sendo ilimitado o número de elementos **meta** cabíveis, desde que cada um contendo seu par **[n][v]** (‘nome’ e ‘valor’).

```
<?xml version='1.0' encoding='utf-8'?>
<document>
  <head id="FL">
    <!--MAP metadata-->
    <metadata generation="MAP">
      <meta>
        <n>codigo_MAP</n>
        <v>0031</v>
      </meta>
      <meta>
        <n>codigo</n>
        <v>97FL</v>
      </meta>
      <meta>
        <n>nome_modernizado</n>
        <v>Francisca Luís</v>
      </meta>
      <meta>
        <n>grafia_conservadora</n>
        <v>Frca Luis; Frca luis; francisa Luis</v>
      </meta>
      <meta>
        <n>trecho_nomeacao</n>
        <v>Francisca Luís, mulher preta, forra, crioula, da cidade do Porto, casada com Domingos Soares, homem pardo, remendão, ausente, do qual não tem novas se é vivo se morto, vendedeira, moradora nesta cidade</v>
      </meta>
    </metadata>
  </head>
</document>
```

Figura 6. Metadados eDictor com categorias MAP

A categoria **metadata** original é preenchida automaticamente com metadados internos, **edictor\_internal** (*data de salvamento, contagem de palavras, etc.*), como ilustra a **Figura 5**. Note-se a estrutura interna básica dos metadados: **[metadata [meta [n v]]]**, onde **metadata** tem um atributo **generation**:

```
<metadata generation="(...)">
  <meta>
    <n>(...)</n>
    <v>(...)</v>
  </meta>
</metadata>
```

Além da instância **metadata** automaticamente gerada, o eDictor permite que se acrescentem ilimitadas outras camadas de **metadata** aos documentos, com qualquer classificação desejada no atributo **generation**<sup>5</sup>.

A **Figura 6** mostra parte do cabeçalho do documento 0031 (Denúncia de Francisca Luís), com a categoria **metadata generation="MAP"** contendo as categorias descritivas do M.A.P. nos seus elementos **meta** (todas as categorias estão registradas no documento; na imagem, são visíveis apenas as cinco primeiras: *códigos, nome modernizado, grafia conservadora do nome e trecho de nomeação*).

Contrastemos isso a um XML básico gerado a partir do Catálogo M.A.P.

O Catálogo conta atualmente com 80 categorias descritivas com uma estrutura plana **[categoria ('conteúdo')]**. A **Figura 7** abaixo mostra isso para o mesmo documento 0031 (sendo visíveis na imagem apenas as primeiras 7 das 80 categorias).

Comparando-se essa estrutura à estrutura mostrada antes, com esses mesmos dados adaptados ao cabeçalho do eDictor, nota-se o caminho da adaptação.

<sup>5</sup> Os metadados do eDictor foram desenvolvidos em 2004, para atender seu primeiro ambiente de uso, o Corpus Tycho Brahe (<https://www.tycho.iel.unicamp.br>), onde trabalhávamos com textos originalmente manuscritos a partir de edições impressas. Queríamos registrar nos XMLs as diferentes informações sobre cada ‘geração’ de editores de um texto; vem daí a ideia das ‘gerações’ de metadados. A ideia era que isso viesse a ser útil também para outros cenários, como de fato aconteceu com os demais projetos que passaram a usar o eDictor.

```
<?xml version='1.0' encoding='utf-8'?>
<registros>
  <catalogada>
    <codigo_MAP>0031</codigo_MAP>
    <codigo>97FL</codigo>
    <nome_modernizado>Francisca Luís
    </nome_modernizado>
    <grafia_conservadora>Frca Luis; Frca luis;
    francisca Luis</grafia_conservadora>
    <trecho_nomeacao>Francisca Luís, mulher
    preta, forra, crioula, da cidade do Porto,
    casada com Domingos Soares, homem pardo,
    remendão, ausente, do qual não tem novas se
    é vivo se morto, vendedeira, moradora nesta
    cidade</trecho_nomeacao>
    <trecho_voz>-</trecho_voz>
    <detalhamento_perfil>Nomeada em processo
    inquisitorial acusada de sodomia feminina</
    detalhamento_perfil>
    <unidade_menor_nascimento>Porto</
    unidade_menor_nascimento>
    <conservadora_menor_nascimento>Salvador
    bahia de todos os Santos; Salvador Bahia
    de todos os Santos; Salvador Capitanja da
    bahia de todos os stos</
    conservadora_menor_nascimento>

```

Figura 7. Metadados do Catálogo M.A.P. (0031)

Importa destacar que o objetivo dessa adaptação de estruturas era integrar as categorias catalográficas **ao mesmo arquivo que contém as edições**. A junção entre o cabeçalho XML do eDictor e o XML do Catálogo M.A.P. cumpre plenamente o fundamento básico do Portal M.A.P., no qual cada instância catalogada corresponde apenas a um documento XML, que inclui os dados catalográficos M.A.P. (na forma dos Metadados eDictor) e a edição filológica do documento.

```
<head id="FL">
<!--MAP metadata-->
  <metadata generation="MAP">...
  </metadata>

<!--edictor metadata-->
  <metadata generation="edictor_internal">
  </metadata>
</head>
<body>
  <text t="full" words="1492" id="text_1">
    title="Denúncia contra Francisca Luís"
    author="Manuel Francisco" year="1592"
    <sc id="sc_1">
      <p id="p_1" t="margin_top_left">
        <s id="s_1">
          <w id="1" t="antropônimo">
            <o>frca</o>
            <e t="exp">francisca</e>
            <e t="atu">Francisca</e>
          </w>
          <w id="2">
            <o>Luis</o>
            <e t="mod">Luís</e>
          </w>
        </s>
      </p>
    </sc>
  </text>
</body>

```

Figura 8. Vista de um XML inteiro, com cabeçalho colapsado e corpo do texto aberto

Note-se então que, para adaptar as categorias M.A.P. aos metadados do eDictor, basta fazer com que os nomes das categorias de catalogação passem a ser o conteúdo do elemento **v** da estrutura do eDictor.

Ou seja: usamos a estrutura **n** e **v** para registrar o nome e valor das categorias catalográficas M.A.P. do documento, como mostra o esquema abaixo:

#### M.A.P. original:

```
<nome_modernizado>Francisca Luís</nome_modernizado>
```

#### M.A.P. eDictor:

```
<meta>
<n>nome_modernizado</n>
<v>Francisca Luís</v>
</meta>
```

Essa junção é explicitada na **Figura 8**, que mostra novo detalhe do documento 0031, agora com as duas ‘gerações’ de metadata em colapso e o ‘texto’ (ou seja, a edição filológica) em proeminência.

Essa estrutura é já bastante satisfatória para nossos objetivos imediatos – entretanto, cabe observar que, no futuro próximo, ela será inteiramente mudada, para adaptar-se ao novo **eDictor 2.0**.

A nova versão da ferramenta inclui um formato de XML compatível com o **TEI (Text Encoding Initiative)**, como discutimos e previmos já em Costa et al. (2022)<sup>6</sup>. A nova versão não foi absorvida pelo M.A.P. até este momento pois só foi terminada em dezembro de 2024, e ainda não está disponível publicamente (cf. 3.1).

Importa notar que o eDictor 2.0 comporta mecanismos simples e automáticos para importação do XML da versão 1.0.

<sup>6</sup> Aline Silva Costa, Aline; Namiuti, Cristiane; Paixão de Sousa, Maria Clara; Costa, Bruno Silvério; Santos, Jorge Viana. An annotation proposal based on TEI Schema for Portuguese Corpora editions: A solution for e-Dictor XML annotation problem. Proceedings of the Second Workshop on Digital Humanities and Natural Language Processing (2nd DHandNLP 2022), 2022. <https://ceur-ws.org/Vol-3128/paper7.pdf>

Assim, mesmo considerando a mudança futura para a nova estrutura de metadados do eDictor, penso que a opção por tomar a estrutura atual como base para o M.A.P. foi apropriada – lembrando que, ao serem importados para o eDictor 2.0, nossos arquivos automaticamente terão seus metadados convertidos para um formato estandardizado internacional.

Considero, por tudo isso, que adaptação dos metadados do Corpus aos metadados do catálogo foi bem-sucedida (ainda que feita ‘ao avesso’, ou seja, adaptamos a estrutura de metadados do Catálogo à estrutura dos metadados do Corpus, e não o posto como prevíamos) e que, apesar de simples, ela é eficiente e flexível, e fica aberta para novas versões em um futuro próximo.

### **2.2.2.3 Adaptação da interface**

Enquanto a adaptação dos metadados se apresenta hoje já muito satisfatória, a adaptação da interface está em um estágio muito incipiente. Na proposta inicial, formulamos da seguinte forma a ideia da integração pela interface:

A direção adequada da integração será do Catálogo para o Corpus: ou seja, **a visualização primária se dará no ambiente que atualmente chamamos de ‘Catálogo’ (i.e., a interface da base de dados)**.

A grande tarefa desse desenvolvimento, portanto, será redesenhar o ambiente visual da base de dados de modo a possibilitar a leitura dos documentos editados (sem que se tenha que sair do ambiente para acessar os textos). (texto do relatório de 2024)

Destaca-se no trecho acima o ponto específico da maior diferença entre os resultados planejados e os efetivamente obtidos: a visualização primária **não está acontecendo a partir da base de dados**, mas sim em um ambiente digital estático, por meio de páginas HTML novas preparadas previamente. Esta não é a implementação ideal concebida na origem do Portal.

Entretanto, a dificuldade enfrentada nesse sentido não é, especificamente, uma consequência da ideia do Portal: é, mais propriamente, uma dificuldade ligada à condição de implementação da base de dados do Projeto de um modo mais geral. É possível dizer que mesmo que tivéssemos mantido a ideia de Catálogo e Corpus separados, o obstáculo teria sido o mesmo. Assim, procuramos agora explicitar resumidamente as razões pelas quais o atraso na implementação do banco impactou a finalização ideal do Portal, e por que teria, de fato, impactado negativamente qualquer outra arquitetura de produto que dependesse do banco ativo.

Começamos pelo final: hoje, os produtos do Projeto estão hospedados em dois servidores da Universidade: o servidor da Pró-reitoria de Pesquisa e Inovação -PRPI, host do nosso site original desde 2017 ([map.prp.usp.br](http://map.prp.usp.br)) e o servidor da Faculdade de Filosofia, Letras e Ciências Humanas - FFLCH ([map.fflch.usp.br](http://map.fflch.usp.br)). O banco de dados está no servidor da FFLCH, e a parte estática do site, no da PRPI. A raiz dessa duplicidade remonta à nossa decisão de migrar o site para o servidor da Faculdade – sobretudo: nossa decisão de lá instalar o novo banco de dados (que estava em fase final de elaboração em meados de 2024). Esse plano teve como fundamento o fato de que, na máquina virtual que temos na PRPI, não nos são oferecidas condições técnicas de operar sistemas dinâmicos – não temos ali segurança, nem apoio institucional. Na FFLCH, ao contrário, contamos com uma equipe de STI dedicada e muito competente. Assim, depois de diversas tentativas infrutíferas de implementar os produtos no prp.usp.br, desistimos, e acordamos com o STI-FFLCH a migração para seu servidor. Esse processo teve início em agosto de 2024 – e, desde então, teria havido tempo hábil para finalizar a migração das páginas estáticas e a instalação dos elementos dinâmicos.

Entretanto, aí nos deparamos com um segundo obstáculo: ao começar a instalação da aplicação Django, do banco de dados e suas funcionalidades no servidor da FFLCH, precisamos adaptar o trabalho já realizado a requerimentos computacionais para os quais nossa equipe

não estava preparada. A responsabilidade disso é, em minha visão, compartilhada: parte nossa, por não ter previsto esses requerimentos, e parte da própria Faculdade, por não apresentar tais requerimentos de modo claro e preliminar aos projetos que precisem fazer uso dinâmico de seus servidores. Salientamos que esse problema não é só nosso, mas também de outros projetos da FFLCH, e que, junto com eles, estivemos procurando soluções institucionais duradoras (comentadas em [2.3.3](#)). Dentre esses requerimentos, citamos alguns que demandaram mais tempo de desenvolvimento, como a necessidade do uso de `cdn` para o `bootstrap`, a mudança do banco de **sqlite** para **MySQL**, a implementação das funções de login e de registro no sistema com requisitos de segurança e o uso do `DotEnv` para as variáveis de ambiente, que necessitaram ser melhor especificadas.

O fato é que esse atraso, alargado de agosto de 2024 até o presente, prejudicou imensamente a finalização da nova interface. Note-se que a interface precisava ser instalada e rigorosamente testada antes de ser considerada pronta; com o atraso, não houve tempo hábil para testes. Esse fato nos desagradou profundamente; restou-nos, entretanto, trabalhar com as condições que estavam colocadas. Assim, tomamos as seguintes providências:

1. Em novembro de 2024, houve reuniões intensivas entre a equipe computacional do Projeto e o STI da FFLCH, delineando um plano para adaptação do banco de dados; o contato dos membros da equipe com o STI se manteve nos últimos meses, sob a capitania da profa. Vanessa Martins do Monte;
2. Como resultado dessa estratégia, o banco de dados foi extensamente reformulado; atualmente, encontra-se em estágio final de adaptação para funcionamento pleno no servidor `fflch.usp.br`. A documentação dessa reestruturação está disponível publicamente no Git-Hub oficial da FFLCH, em [github.com/fflch/map](https://github.com/fflch/map) ;
3. Por ora, o acesso externo ao nosso site no servidor da FFLCH apresenta funcionalidades limitadas, como se pode ver em [map.fflch.usp.br](https://map.fflch.usp.br). Está ativa ali a página de acesso para catalogação, mediante senha; mas não a interface para consulta dos usuários;
4. Como medida paliativa, para consulta externa compusemos páginas HTML que aproveitam o conteúdo da base de dados de forma estática (ou seja, gerando XMLs a partir do banco e gerando localmente HTMLs a partir deles, e posteriormente fazendo upload desses HTMLs para o servidor da PRPI). Isso resulta na interface provisória apresentada brevemente em [2.2.1.2](#) acima, e disponível a partir de [map.prp.usp.br/MAP.html](https://map.prp.usp.br/MAP.html).

A situação está longe do ideal; entretanto, é razoavelmente funcional e, esperamos, provisória. Em [2.2.3](#) adiante pontuamos as lições aprendidas e o planejamento frente a esse problema.

### 2.2.2.2 Preparo de documentação

Conforme previmos na proposta de reestruturação dos produtos do Projeto, o aumento na complexidade das técnicas computacionais e filológicas desenvolvidas traria a necessidade de oferecer uma documentação detalhada que integrasse desde as decisões pertinentes ao trabalho de edição até a organização digital da base documental. Planejamos que essa documentação, o '**Manual M.A.P.**', fosse publicada em conjunto com o Portal, destinando-se tanto a futuros pesquisadores como ao usuário final do sistema, e composto de dois documentos técnicos – Manual de Catalogação e Manual de Edições Filológicas Digitais.

Estão hoje publicados o **Manual de Edição Filológica** e o **Manual de Descrição Documental**, cf. [map.prp.usp.br/MAP\\_Manual.html](https://map.prp.usp.br/MAP_Manual.html). A documentação inclui aspectos conceituais e técnicos de nossas opções metodológicas de curadoria e edição.

Apresentação .....	3
1 Catálogo Digital M.A.P. ....	4
1.1 Catalogação .....	6
1.2 Inserção de documento .....	7
1.3 Visualização das catalogadas .....	9
1.4 Visualização dos documentos .....	11
2 Categorias .....	13
2.1 Definição e Estrutura das Categorias .....	13
2.2 Grupos de Categorias .....	13
2.2.1 Informações sobre a catalogada .....	14
2.2.2 Informações sobre o documento .....	19

Figura 9. Sumário do Manual de Descrição Documental.

Apresentação .....	4
1. Processo de edição de texto .....	6
1.0 Funcionamento básico do eDictor.....	6
1.1 Edição de texto e etiquetas.....	8
1.2 Descrição de elementos de texto .....	14
1.3 Metadados .....	29
1.4 Apresentação final das edições .....	33
2. Normas de edição .....	36
2.0 Etapa de Transcrição.....	36
2.1 Camada diplomática .....	39
2.2 Camada semidiplomática .....	41
2.3 Camada atualizada .....	44
2.4 Camada modernizada.....	47
2.5 Versões para visualização .....	49
4. Como referenciar.....	51
4.1 O Corpus M.A.P.....	51
4.2 As edições eletrônicas .....	51
eDictor: tutorial de instalação e customização ....	52
1.1 Download e instalação .....	53
1.2. Módulo de transcrição .....	55
1.3. Módulo de edição .....	58

Figura 9. Sumário do Manual de Edição Filológicas

A documentação publicada, como um todo, atendeu aos requisitos do nosso plano original; mas o Manual de Edições superou as expectativas iniciais. O envolvimento da equipe filológica nesse trabalho nos meses finais da vigência do fomento foi notável. Em particular, isso se intensifica a partir de novembro de 2024, ocasião da **Oficina de Produtos M.A.P.**, quando as bolsistas se debruçaram de modo minucioso sobre as diferentes opções de edição ocasionadas pelas distintas tipologias documentais estudadas por cada uma. Foi o caso, em especial, das decisões e desafios em torno das abreviaturas e da descrição diplomática dos documentos. Esse rico trabalho trouxe para as coordenadoras do Projeto a ideia de as bolsistas redigirem um artigo futuro sobre o Manual de Edição e as decisões e discussões conceituais que ele proporcionou.

O **Manual de Descrição Documental** começa com uma breve apresentação da concepção e dos princípios metodológicos básicos do trabalho de curadoria no Projeto M.A.P.. A seção 1 descreve as técnicas básicas para catalogar os documentos na base. A seção 2 apresenta um panorama completo das 80 categorias descritivas, explicando-as e trazendo exemplos. A metodologia descrita neste Manual foi discutida por toda a equipe em múltiplas reuniões, e sua elaboração final ficou a cargo de Priscila Tuy e Igor Leal.

Este Manual está disponível para download em [map.prp.usp.br/Manual/MAP\\_descricao.pdf](http://map.prp.usp.br/Manual/MAP_descricao.pdf)

O **Manual de Edição Filológica** tem início com uma apresentação sobre a concepção de edições filológicas eletrônicas seguidas no Projeto M.A.P.. A seção 1 apresenta as técnicas fundamentais de uso do eDictor para edição de textos e explica o uso personalizado concebido pela equipe M.A.P, com profusa exemplificação. A seção 2 apresenta as normas de edição seguidas no Projeto, em todas as camadas e etapas do trabalho. O manual inclui ainda um apêndice com um breve tutorial para instalação do eDictor.

As técnicas e normas apresentadas no Manual foram desenvolvidas pela equipe filológica em diversas reuniões, oficinas e seminários. A elaboração final do documento ficou a cargo de Andréa Natanael, Elisa Motta e Nicólli Garcia, com apoio de Priscila Tuy e revisão final das coordenadoras do Projeto.

Este Manual está disponível para download em [map.prp.usp.br/Manual/MAP\\_edicao.pdf](http://map.prp.usp.br/Manual/MAP_edicao.pdf)

## 2.3 Outras ações, produtos e resultados

### 2.3.1 O eDictor 2.0

O desenvolvimento do **eDictor 2.0** esteve a cargo de **Aline Silva Costa**, pesquisadora não bolsista deste projeto, conforme nossa proposta original, e como parte da tese de doutoramento de Aline, conduzida na **Universidade Estadual do Sudoeste da Bahia (UESB)**, orientada por **Cristiane Namiuti**, também pesquisadora associada ao Projeto, e coorientada por mim (Costa, 2024)<sup>7</sup>. Por conta de um adiamento na conclusão da tese, o eDictor 2.0 ainda não estava pronto por ocasião do relatório de maio de 2024. Em dezembro, a tese foi defendida, com sucesso. O trabalho de Aline em torno da modernização e atualização dessa ferramenta pode ser chamado de exemplar, como pude acompanhar nas sessões de coorientação e tendo em vista a reação positiva da banca do exame. Volto a destacar aqui, como fiz no relatório anterior, a complexidade envolvida no trabalho: não se tratou apenas de produzir um eDictor superficialmente distinto, e sim de desenvolver um sistema de anotação inteiramente renovado, em conformidade com padrões internacionais. Considero o resultado excelente e, como desenvolvedora do eDictor original, julgo esse um passo feliz avanço, a quase vinte anos do lançamento do eDictor Beta.

Entretanto, há um problema sério: embora esteja pronta, a ferramenta ainda não está disponível publicamente; de fato, não pode ser usada por nenhum pesquisador fora do ambiente interno do servidor da UESB. Malgrados repetidos esforços, essa contingência não foi resolvida até o presente. Dessa forma, ao contrário do que havíamos planejado, não foi possível para a equipe de bolsistas do M.A.P. testar o novo aplicativo (o que desejávamos fazer ainda em 2024, tendo em vista que nossa intenção sempre foi migrar para o eDictor 2.0). Esse objetivo não cumprido fica reservado para projetos futuros. Nesse sentido importa ressaltar que uma das vantagens da metodologia de Costa (2024) é a robusta interoperabilidade do novo sistema, inclusive no sentido de poder importar os produtos do eDictor antigo e adaptá-lo integralmente ao novo ambiente e ao novo esquema de anotação. Assim, todo o trabalho feito por nós até agora no eDictor 1.0 poderá ser aproveitado quando elevarmos o Corpus para a ferramenta mais nova.

### 2.3.2 A pesquisa em torno do Handwritten Text Recognition (HTR)

A pesquisa e experimentação em tecnologias de reconhecimento automático de manuscritos (HTR) foi parte importante da proposta inicial do Projeto. Seus resultados não figuraram com destaque neste relatório nem no relatório de 2024 uma vez que essa área dos trabalhos foi adiantada para o primeiro ano da vigência do fomento, como relatado em 2023. Destacamos aqui os resultados principais então narrados, acrescentados de dois acontecimentos recentes.

Como resultados já relatados, lembramos a intensa pesquisa com a leitura automática dos softwares **LapeLinc** e **Transkribus** em um subcorpus de nove documentos do Corpus M.A.P, com excelentes resultados, discutidos no artigo recentemente aceito para publicação no periódico **Digital Scholarship in the Humanities**, da Universidade de Oxford, veículo de grande impacto e prestígio no campo internacional das Humanidades Digitais (cf. 1.3).

O sucesso nos experimentos gerou também grande interesse nosso na continuidade do uso do software Transkribus, mantido pela cooperativa **ReadCoop**, da Universidade de Innsbruck, Áustria. Trata-se de um software de HTR único em escala internacional, utilizado por universidades do mundo todo, que contribuem com o desenvolvimento de modelos para leitura

<sup>7</sup> Costa, Aline Silva. Um sistema de anotação de múltiplas camadas para Corpora (Históricos) de Língua Natural. Tese de doutoramento. Programa de Pós-graduação em Linguística, Universidade Estadual do Sudoeste da Bahia, 13 de dezembro, 2024.

dos mais variados tipos caligráficos. Nossas pesquisas até este ponto deixaram evidentes as potencialidades do software para a leitura ‘automática’ de alguns dos manuscritos do Catálogo M.A.P. – em particular, os documentos inquisitoriais. Assim, no segundo semestre de 2024, mantivemos contato com o representante da ReadCoop para a América Latina, Sr. Eugenio Torres, que sugeriu que nosso grupo se tornasse membro da cooperativa. Com isso, a USP seria a primeira universidade Brasileira a fazer parte da iniciativa. As tratativas com o ReadCoop avançaram bastante; infelizmente, entretanto, não conseguimos fazer o pagamento das ‘cotas’ (*‘shares’*) com os recursos Fapesp – pois, apesar de agência ter aprovado a despesa, não houve tempo hábil para a conversão dos valores de dólares a reais, uma vez que as tratativas se estenderam e a liberação final aconteceu depois dos 30 dias anteriores ao término do prazo de vigência. Entretanto, não queremos perder essa oportunidade e contato; assim, para um próximo projeto de fomento, utilizaremos a forma de contratação já aceita pela Fapesp.

### **2.3.3 Observações sobre o apoio institucional à pesquisa**

Comentamos, em **2.3.2** acima, as dificuldades enfrentadas pelo Projeto na implementação de nossos recursos tecnológicos de forma ativa em servidores da USP; comentamos também que o problema não se coloca apenas para nosso Projeto, e sim é geral em nossa Faculdade. Diante disso, além de buscar remédios paliativos para a finalização de nossos produtos e das diversas reuniões e tratativas com o STI da FFLCH, estivemos também procurando soluções institucionais mais duradouras, pensando em pesquisas futuras.

Nesse sentido, estivemos reunidas com a direção da Faculdade, em conjunto com coordenadores de outros projetos apoiados pela Fapesp, para organizar a iniciativa de se formar um laboratório de apoio tecnológico voltado especialmente aos projetos com uso intensivo de tecnologias digitais – talvez, de fato, um Laboratório de Humanidades Digitais da Faculdade.

Esse espaço de apoio poderia funcionar nos moldes do Escritório de Apoio Institucional ao Pesquisador (EAIP, [eaip.fflch.usp.br](http://eaip.fflch.usp.br)). Como pudemos constatar ao longo da vigência deste Projeto, o EAIP é hoje um ponto de apoio institucional absolutamente fundamental para o bom transcorrer das operações relativas a compras e organização financeira. A competência de seus funcionários é admirável, e a existência do espaço dedicado aos Projetos facilita imensamente a interação entre os projetos e eles. Esperamos que a ideia do novo laboratório também dê frutos, e que a sua instauração possa facilitar a interação entre os projetos e o excelente time do STI.

## **3 Fechamento e perspectivas**

Em seus sete anos de trabalhos, o Projeto **M.A.P.: Mulheres na América Portuguesa** acumulou uma experiência considerável no estudo da documentação sobre as mulheres na América Portuguesa de uma perspectiva filológica orientada por métodos computacionais. Iniciado em 2017, com o objetivo de sistematizar e tornar visíveis para pesquisas futuras documentos escritos por mulheres e sobre mulheres na América Portuguesa entre 1500 e 1822 a partir da perspectiva filológica e com base em métodos das Humanidades Digitais, o Projeto logrou reunir um importante conjunto documental pertinente à história das mulheres no Brasil entre os séculos XVI e XIX, com relevância para os estudos filológicos e para os estudos da história da língua, da história social, da história da escrita e da leitura, e da história das mulheres no Brasil.

Vemos nosso Projeto como um empreendimento de longo prazo, composto de ciclos com metas específicas que se sobrepõem aos ciclos anteriores aproveitando e expandindo seus resultados.

Em sucessivos ciclos focais e com diferentes estratégias de fomento, o Projeto produziu o **Catálogo M.A.P.** (lançado em 2018), o **Corpus M.A.P.** (iniciado em 2021) e o **Portal M.A.P.** (lançado em janeiro de 2025). O Portal M.A.P. reúne hoje Catálogo e Corpus, e configura o resultado principal deste ciclo de fomento Fapesp entre 2021 e 2025. No encerramento deste ciclo, já idealizamos as novas etapas de trabalhos no M.A.P. a serem iniciadas neste ano de 2025.

Pretendemos trazer a foco agora alguns dos desafios e potencialidades importantes revelados pelos ciclos anteriores, sempre buscando o adensamento e o refinamento de nossa compreensão sobre as questões mais amplas do Projeto. Nossa perspectiva central na nova etapa será aprofundar o estudo paleográfico, diplomático e codicológico dos manuscritos relevantes para a pesquisa, traduzindo esse aprofundamento em novos produtos representativos do estado-da-arte das tecnologias computacionais voltadas para as Humanidades. Tornou-se claro para nós, neste ponto da história do Projeto, o potencial encerrado nos documentos que formam nosso Portal para diferentes áreas de estudo e, particularmente, o potencial de alguns deles para o tratamento computacional sofisticado. A esses documentos queremos nos dedicar no futuro próximo.



Tribunal do Santo Ofício (TSO). Repertório geral das mulheres, 1º tomo. Inquisição de Lisboa, liv. 28. 1143 f. Foto tirada por mim figurando a prof.ª Vanessa; julho, 2024.

Esse novo horizonte para as pesquisas vinha se desenhando gradualmente conforme se aproximava o final da vigência do presente projeto; entretanto, as perspectivas ficaram particularmente claras como resultado da viagem de campo realizada em julho de 2024 junto ao **Arquivo Nacional da Torre do Tombo**, em Lisboa. Na ocasião, as coordenadoras do Projeto estudaram minuciosamente o primeiro tomo do Repertório Geral das Mulheres (cf. 2.1.2).

É difícil expressar o impacto e o potencial de estudo futuro deste material sem recorrer a expressões hiperbólicas. Em seus 1.143 fólios, o livro apresenta uma indexação interna extremamente complexa, com imensa riqueza codicológica, diplomática e paleográfica.

Na presença do livro e examinando seus intrincados mecanismos de referência, imediatamente imaginei uma indexação digital que permitisse não apenas a pesquisa ágil do volume como sua ligação com os documentos relativos aos casos citados. Editar digitalmente um material com este perfil e extensão, entretanto, não é um empreendimento trivial. Por isso, meu segundo pensamento foi o de como as tecnologias de HTR seriam ideais para essa tarefa.

Surge assim a ideia de uma nova proposta de fomento, que será enviada à Fapesp ainda neste primeiro semestre de 2025, e que se concentrará no material indexatório produzido pelo Tribunal do Santo Ofício. Pretendemos desenvolver uma metodologia que permita traduzir a indexação orgânica encerrada em manuscritos em indexações digitais. Sobretudo, queremos criar as condições para a articulação entre o aprofundamento do trabalho filológico e a intensificação do desenvolvimento computacional no Projeto, em particular no que toca as tecnologias de HTR.

**O Corpus M.A.P.:  
Desenvolvimento de um sistema de  
edições filológicas digitais para a  
documentação sobre as Mulheres na  
América Portuguesa**

**Relatório Científico  
Fevereiro, 2025**

**Anexos**

**Maria Clara Paixão de Sousa  
Universidade de São Paulo**

# **Lista de anexos**

## **Anexo 1:**

Relatório Sintético do bolsista  
**Igor Leal Souza**

## **Anexo 2:**

Relatório Sintético da bolsista  
**Priscila Starline Estrela Tuy Batista**

## **Anexo 3:**

Relatório Sintético da bolsista  
**Elisa Hardt Leitão Motta**

## **Anexo 4:**

Relatório Sintético da bolsista  
**Andréa Cristina Natanael da Silva**

## **Anexo 5:**

Relatório Sintético da bolsista  
**Nicólli de Lima Garcia**

## **Anexo 1:**

**Relatório Sintético do bolsista**  
**Igor Leal Souza**

**Corpus M.A.P.:**

**Desenvolvimento do Catálogo M.A.P. 2.0 e dos  
metadados para o Corpus M.A.P. Beta  
(2022/16216-1)**

## **Anexo 2:**

**Relatório Sintético da bolsista  
Priscila Starline Estrela Tuy Batista**

**Corpus M.A.P.:**

**Preparo dos fundamentos e da visualização final  
de manuscritos com edição filológica digital  
(2022/16217-8)**

## **Anexo 3:**

**Relatório Sintético da bolsista  
Elisa Hardt Leitão Motta**

**O Corpus M.A.P.:**

**Edição e coordenação da revisão da edição  
filológica digital dos documentos provenientes  
do Arquivo Público do Estado de São Paulo  
(APESP)  
(2022/16215-5)**

## **Anexo 4:**

**Relatório Sintético da bolsista  
Andréa Cristina Natanael da Silva**

**O Corpus M.A.P.:  
Edição e coordenação da revisão da edição  
filológica digital dos processos inquisitoriais  
provenientes do Arquivo Nacional da Torre do  
Tombo (ANTT)  
(2022/16337-3)**

## **Anexo 5:**

**Relatório Sintético da bolsista  
Nicólli de Lima Garcia**

**O Corpus M.A.P.:  
Edição fac-similar dos documentos provenientes  
do Arquivo Público do Estado de São Paulo  
(APESP) e edição filológica digital dos  
manuscritos do Instituto de Estudos Brasileiros  
da USP (IEB)  
(2023/06772-7)**