
Plan Overview

A Data Management Plan created using DMPTool

Title: Projeto M.A.P. (Mulheres na América Portuguesa)

Creator: Maria Clara Paixão de Sousa - **ORCID:** [0000-0002-8422-417X](https://orcid.org/0000-0002-8422-417X)

Affiliation: Universidade de São Paulo (www5.usp.br)

DMP ID: <https://doi.org/10.48321/D1801T>

Funder: São Paulo Research Foundation (fapesp.br)

Template: Digital Curation Centre (português)

Project abstract:

O objetivo desta proposta é dar início à formação do **Corpus M.A.P.**, uma reunião inédita de manuscritos de mulheres e sobre mulheres da América Portuguesa editados filologicamente com tecnologias digitais de última geração. As fontes do Corpus integram o **Catálogo M.A.P.**, repositório digital criado no âmbito do **Projeto M.A.P. (Mulheres na América Portuguesa)**. O projeto, ativo desde 2017, tem como objetivo central sistematizar e tornar visível para pesquisas futuras um conjunto de fontes documentais imensamente importantes 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, por meio da construção de um catálogo eletrônico de documentos escritos por mulheres na América Portuguesa entre 1500 e 1822. A relevância do Projeto reside fundamentalmente na possibilidade de organização inédita dessa documentação a um tempo escassa e fundamental para a compreensão da história da formação do Brasil. Presentemente, o Catálogo comporta 150 entradas detalhadamente descritas e um texto-piloto com edição filológica digital. A sistematização integral dessa documentação na forma do Corpus aqui proposto permitirá a ampliação do alcance do público leitor dos textos e a disponibilização desse material para fins de pesquisa nos campos da linguística e da ciência da computação. A proposta se filia aos trabalhos do Grupo de Pesquisas Humanidades Digitais (formado por pesquisadoras do M.A.P) junto ao Centro de Inteligência Artificial (C4AI) da Universidade de São Paulo, atualmente responsáveis pela formação do Corpus Carolina (com lançamento previsto para março de 2022).

Start date: 08-01-2022

End date: 07-01-2024

Last modified: 12-23-2021

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Projeto M.A.P. (Mulheres na América Portuguesa)

Coleta de Dados

Que dados serão coletados ou criados?

A meta do projeto é 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, organizados em três conjuntos de acordo com suas distintas tipologias documentais de origem:

1. Documentos do Fundo do Tribunal do Santo Ofício (TSO): 10
2. Documentos do Fundo do Governo da Capitania de São Paulo (GCSP): 20
3. Documentos do Arquivo do Instituto de Estudos Brasileiros da USP (IEB): 10

A maior parte desses documentos (conjuntos A e B) já estão catalogados e descritos pela equipe do Projeto, em <http://map.prp.usp.br/MAP=Fichas.html>; essa catalogação inclui a atribuição de créditos e ligações para imagens facsimilares dos documentos digitalizados em seus arquivos de origem, quando aplicável. O principal dado novo a ser criado no Projeto será o conjunto formado pelas edições digitais desses documentos, seguindo os princípios da prática filológica.

As edições digitais a serem produzidas seguirão, em larga medida, o modelo da edição-piloto publicada em <http://map.prp.usp.br/Corpus/FL/FL.html>. Incluirão, assim, a reprodução dos facsímiles, o XML de base da edição, e diferentes versões HMTL (edição conservadora diplomática e edição modernizada). O acesso aos documentos será aberto, livre e gratuito.

Como os dados serão coletados ou criados?

Documentação e Metadados

Que documentação e metadados acompanharão os dados?

O projeto pretende codificar o corpus na linguagem XML, de acordo com os padrões W3, em formato aberto e livre. Os metadados do corpus seguirão, ainda, os padrões TEI (cf. TEI Consortium, eds. TEI P5: Guidelines for Electronic Text Encoding and Interchange. [Version 4.3.0.]. [Last updated 31st August 2021]. TEI Consortium. <http://www.tei-c.org/Guidelines/P5>.)

Ética e Conformidade Legal

Como você administrará qualquer questão ética?

Os documentos a serem editados no Corpus são provenientes de arquivos públicos, físicos e digitais:

1. Arquivo Nacional da Torre do Tombo (ANTT) - Fundo do Tribunal do Santo Ofício (TSO)
2. Arquivo Público do Estado de São Paulo (APESP) - Fundo do Governo da Capitania de São Paulo (GCSP)
3. Arquivo do Instituto de Estudos Brasileiros da USP (IEB)

Todos os créditos aos arquivos de origem serão concedidos. No caso dos documentos prospectados em arquivos físicos (até o momento, no Arquivo Público do Estado de São Paulo e no Arquivo do Instituto de Estudos Brasileiros da Universidade de São Paulo), obtivemos licença para fotografar os documentos e para estudo e divulgação do material.

Como você vai gerenciar os direitos autorais e os direitos de propriedade intelectual (IP / IPR)?

Essa pesquisa não envolverá fontes originais com incidência de direitos autorais, tendo em vista o escopo temporal do material a ser editado no corpus (manuscritos produzidos entre 1500 e 1822). Assim, todos os dados documentais utilizados na pesquisa são de domínio público, de acordo com a Lei do Direito Autoral (Lei nº 9.619/98) que prevê que os direitos patrimoniais das obras perduram até 70 anos após o falecimento do autor.

Armazenamento e Backup

Como os dados serão armazenados e terão backup durante a pesquisa?

Os dados a serem produzidos serão publicados no site do projeto, em <http://map.prp.usp.br>. O site tem como base uma máquina virtual dedicada no sistema Internuens da Universidade de São Paulo. O projeto tem à disposição nessa máquina 1TB de espaço para armazenamento.

Como você vai gerenciar o acesso e a segurança?

Question not answered.

Seleção e Preservação

Quais dados são de valor a longo prazo e devem ser mantidos, compartilhados e / ou preservados?

Os dados criados pelo projeto na forma de documentos editados filologicamente terão valor a longo prazo para pesquisas em diversos campos - em particular, 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 acesso aos documentos será aberto, livre e gratuito, e sua codificação será preparada em linguagens de anotação também abertas e livres.

Qual é o plano de preservação a longo prazo do conjunto de dados?

Question not answered.

Compartilhamento de Dados

Como você vai compartilhar os dados?

A exemplo dos dados já preliminarmente trabalhados no Projeto, pretendemos publicar esses dados com licença do tipo Creative Commons - Atribuição-NãoComercial 4.0 Internacional. Desse modo, os usuários poderão compartilhar e distribuir o material, e transformar e criar a partir do material, concedendo o crédito apropriado e indicando se mudanças foram feitas, mas sem sugerir que o licenciante apoia esse uso. O material não poderá ser usado para fins comerciais.

Existem restrições ao compartilhamento de dados requeridos?

Question not answered.

Responsabilidades e Recursos

Quem será responsável pelo gerenciamento de dados?

Question not answered.

Quais recursos você precisará para entregar seu plano?

Question not answered.

Planned Research Outputs

Dataset - "Corpus M.A.P. Beta"

O Corpus M.A.P Beta é 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 Beta será composto por 40 documentos manuscritos, provenientes do Fundo do Tribunal do Santo Ofício do Arquivo Nacional da Torre do Tombo, do Fundo do Governo da Capitania de São Paulo do Arquivo Público do Estado de São Paulo e do Arquivo do Instituto de Estudos Brasileiros da USP (IEB):

10

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Corpus M.A.P. Beta	Dataset	2024-01-31	Open	None specified		Creative Commons Attribution Non Commercial 4.0 International	TEI (Text Encoding Initiative) Guidelines	No	No