

## PLANO DE GESTÃO DE DADOS (PGD)

### 1 Coleção de dados

#### 1.1 Que dados serão coletados ou criados?

Orientação:

- **Descrição dos Dados** – Descrever resumidamente os dados a serem coletados, produzidos ou reutilizados para a investigação da hipótese de pesquisa formulada. Justifique sua escolha de formato e considere as implicações do formato e dos volumes de dados em termos de armazenamento, *backup* e acesso.
- **Perguntas orientadoras:**
  - ✓ Qual o tipo, o formato e o volume dos dados (por exemplo: dados tabulares em planilhas Excel com ocupação de espaço de aproximadamente 10 Mbytes)?
  - ✓ Os formatos e software escolhidos permitem o compartilhamento e o acesso de longo prazo aos dados?
  - ✓ Há dados possíveis de serem reutilizados?

Preencheruvycycycycycycycycyccyvvyvvv aqui:

#### 1.1 Descrição dos dados

##### **Tipo, formato e volume quanto dos dados coletados:**

Mapa Bibliográfico de Gramsci no Brasil - pesquisa realizada em 2016 (1ª ed.), 2018 (2ª ed.) e 2019 (3ª ed.) pela equipe do Núcleo de Estudos e Pesquisas em Filosofia, Política e Educação (NuFiPE/UFF). Trata-se do levantamento bibliográfico das obras que, no Brasil, discorrem acerca de Antonio Gramsci como referencial teórico e/ou a respeito de sua biografia.

O material é divulgado em formato textual, em um arquivo de tipo PDF, disponibilizado online na página da International Gramsci Society-Brasil (<https://storage.googleapis.com/production-hostgator-brasil-v1-0-9/739/971739/23GwhSwA/93b61e567faa4c7d906c5568c73acb77?fileName=IGS-Mapa->

Bibliogr%C3%A1fico-de-Gramsci-no-Brasil-Site-Setembro-de-2019-1.pdf), compondo um volume de aproximadamente 2 mb.

### **Quanto aos dados a serem criados:**

A base coletada em formato PDF será transformada em um dado estruturado (em formato CSV, JSON-LD e RDF), seguindo os princípios FAIR, e disponibilizada em repositório com identificador persistente (DOI) e controle de versionamento (utilizando o repositório do projeto no Zenodo associado ao GitHub), contendo marcação com metadados de origem e proveniência de modo a assegurar informações sobre o ciclo de vida dos dados de pesquisa e a relação com a organização e a comunidade responsável pelo conteúdo ([referência: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009041](https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009041)). Além disso, serão adicionados metadados descritivos sobre o conteúdo das publicações para utilização no grafo de conhecimento a ser produzido.

### **- Os formatos e software escolhidos permitem o compartilhamento e o acesso de longo prazo aos dados?**

Os dados coletados dispõem de um endereço (URL), mas ainda não dispõem de identificadores persistentes, controle de versionamento e metadados de proveniência e autoridade acionáveis por máquinas, como preconizado pelas boas práticas para dados na web, publicadas como recomendação pela W3C (2019 - <https://www.w3.org/Translations/DWBP-pt-BR/>).

) e pelos princípios de dados FAIR, conforme indicado por Wilkinson (2016). (Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>).

Em razão dessas lacunas, a pesquisa pretende transformar esses dados adotando um modelo de publicação compatível com os princípios FAIR, para os dados e metadados a serem trabalhados.

Os dados e as transformações realizadas serão organizados por meio de um script em Python, utilizando as bibliotecas Requests e Pandas, e divulgados em um Caderno Jupyter nos repositórios do projeto (Zenodo e GitHub), facilitando a reprodutibilidade da pesquisa e o acesso de longo prazo.

### **- Há dados possíveis de serem reutilizados?**

Os dados coletados podem ser reutilizados uma vez que foram publicados pela IGS-Brasil a qual tem, dentre suas finalidades estabelecidas em estatuto, [Art. 2.º A IGS-Brasil tem por finalidades: I - Promover a criação de um marco institucional mediante o qual indivíduos e grupos intelectuais, movimentos sociais e político, outras associações, instituições, organizações e entidades do Brasil, interessados na vida e no pensamento de Antonio Gramsci e nos múltiplos problemas que derivam da sua obra, possam intercambiar informações relacionadas às suas pesquisas, aos seus projetos editoriais, às suas particulares perspectivas analíticas e suas atividades em geral dentro do universo do pensamento gramsciano;](#)

V - Envidar esforços para garantir o acesso do público brasileiro à bibliografia gramsciana publicada em âmbito nacional e internacional;

(<https://storage.googleapis.com/production-hostgator-brasil-v1-0-9/739/971739/23GwhSwA/8c29ddaa51774557afaeab3af18f4732?fileName=Estatuto%20IGS.pdf>)

Os dados criados poderão ser reutilizados conforme a licença estabelecida para todos os conteúdos produzidos no âmbito desse projeto (a saber, CC-BY 4.0).

## 1.2 Como os dados serão coletados ou criados?

Orientação:

- **Obtenção dos Dados** – Relatar resumidamente como serão obtidos os dados coletados, produzidos ou reutilizados. Pode incluir processos como repetição de amostras ou medições, captura ou registro de dados padronizados, validação de entrada de dados, revisão por pares de dados ou representação com vocabulários controlados.
- **Perguntas orientadoras:**
  - ✓ Quais padrões ou metodologias serão usados?
  - ✓ Como serão estruturados e nomeados as pastas e arquivos?
  - ✓ Como se lidará com o controle de versão?
  - ✓ Que processos de garantia de qualidade serão adotados?

Preencher aqui:

### 1.2 Obtenção dos dados coletados:

Os dados serão coletados e criados utilizando um script Python a ser disponibilizado por meio de um Caderno Jupyter, facilitando a reprodutibilidade da pesquisa e a reutilização dos dados.

Será realizado download do arquivo PDF da URL <https://storage.googleapis.com/production-hostgator-brasil-v1-0-9/739/971739/23GwhSwA/93b61e567faa4c7d906c5568c73acb77?fileName=IGS-Mapa-Bibliogr%C3%A1fico-de-Gramsci-no-Brasil-Site-Setembro-de-2019-1.pdf>, e se seguirão as transformações necessárias para tornar os dados textuais brutos em dados estruturados e marcados com metadados adequados.

As pastas e arquivos serão nomeadas em formato de *slug* (texto em letras minúsculas separados por hífen), visando a fácil previsibilidade de URL/URI nos repositórios do projeto no GitHub e no Zenodo. Os arquivos dentro dos repositórios serão separados em pastas destinadas aos: arquivos do projeto (projeto), dados coletados (dados-coletados), Caderno Jupyter contendo os scripts em Python (caderno-jupyter), dados criados (dados-criados) e publicações (publicacoes).

- <https://github.com/emiliefaedo/gramsci-kg>

O controle de versão será realizado por meio dos repositórios do GitHub e do Zenodo que já estão associados e dispõem de sistemas robustos de versionamento. Será utilizado o modelo de versionamento semântico (Semantic Versioning 2.0.0 - <https://semver.org/>) como critério para definir as versões dos arquivos.

O processo de garantia de qualidade contará com a verificação de tipos de dados em Python, realizada por meio da biblioteca em Pydantic, a qual permite a validação dos dados de entrada e fornece um conjunto de validadores predefinidos para vários tipos de dados, incluindo strings, números e datas. Além de possibilitar a definição de validadores personalizados usando instrução interna. O processo também será documentado no Caderno Jupyter do projeto.

## 2 Documentação e metadados

### 2.1 Que documentação e metadados irão acompanhar os dados?

Orientação:

- **Metadados** – Indicar os metadados e os padrões adotados para ajudar os usuários secundários a entendê-los e reutilizá-los. Isso deve incluir pelo menos detalhes básicos que ajudarão as pessoas a encontrar os dados, incluindo encontrar quem os criou ou contribuiu com os dados, seu título, data de criação e em que condições eles podem ser acessados. A documentação

também pode incluir detalhes sobre a metodologia usada, informações analíticas e procedimentais, definições de variáveis, vocabulários, unidades de medida, quaisquer suposições feitas e o formato e tipo de arquivo dos dados.

- Perguntas orientadoras:

- ✓ Que informações são necessárias para que os dados sejam lidos e interpretados no futuro?
- ✓ Como será capturada/criada a documentação e os metadados?
- ✓ Quais padrões de metadados você usará e por quê?
- ✓ Os dados de pesquisa se adequam ao princípio FAIR (Localizável, Acessível, Interoperável, Reutilizável)?

Preencher aqui:

Será utilizado como padrão para os metadados um subconjunto dos vocabulários da Dublin Core (<http://purl.org/dc/terms/>), adequados à descrição de cada recurso bibliográfico disponibilizado no arquivo coletado. No desenvolvimento do trabalho, caso seja identificada a necessidade de metadados adicionais, será dada preferência aos conjuntos de metadados disponibilizados como recomendação pela W3C e aos identificadores publicados na Wikidata, no intuito de facilitar a compreensão e o reuso dos dados em novas pesquisas.

Os metadados serão divulgados dentro dos objetos digitais gerados durante o projeto com a base de dados tratada para a elaboração do grafo de conhecimento, em formato CSV, JSON-LD e RDF, com identificadores persistentes e metadados de proveniência e autoridade, alinhados com os princípios FAIR.

Destaca-se que os vocabulários de metadados do Dublin Core, das recomendações da W3C (PROV-O, SKOS, RDF, JSON-LD e outros) e da Wikidata contam com vasta utilização e documentação por uma comunidade especializada na criação e curadoria de descritores, o que tende a facilitar com que os dados criados no âmbito deste projeto sejam lidos e interpretados no futuro.

### 3 Ética e conformidade legal

#### 3.1 Como serão gerenciadas quaisquer questões éticas?

Orientação:

- **Aspectos Éticos e Legais** – Relatar como serão observados os aspectos legais, éticos e de conduta, indicando como serão assegurados o cumprimento das leis de proteção de dados e de propriedade intelectual.
- Perguntas orientadoras:
  - ✓ Há consentimento para preservação e compartilhamento de dados?
  - ✓ Como será protegida a identidade dos participantes, se necessário? (exemplo: via anonimização)
  - ✓ Como os dados confidenciais serão tratados para garantir que sejam armazenados e transferidos com segurança?

Preencher aqui:

### 3. Ética e conformidade legal

Os dados coletados estão publicados em acesso público, não haverá a necessidade de proteger a identidade de participantes, nem serão utilizados dados confidenciais. Os dados são de autoria da equipe do Núcleo de Estudos e Pesquisas em Filosofia, Política e Educação (NuFiPE/UFF) e estão publicados pela IGS-Brasil.

Quanto aos direitos autorais, os dados criados serão disponibilizados com licença que permite seu reuso para qualquer fim, desde que citada a fonte, uma vez que será utilizado para licenciamento de todos os conteúdos produzidos ao longo da pesquisa a licença **Creative Commons Atribuição 4.0 Internacional (CC-BY 4.0)**. Esta licença permite que o trabalho seja copiado e redistribuído em qualquer suporte ou formato. Também permite que o trabalho seja adaptado, re combinado, transformado a partir do material produzido, para qualquer fim, mesmo que comercial. Tais características de licenciamento são compatíveis com os princípios de ciência aberta.

Não haverá período de embargo e restrição para os dados criados uma vez que o conteúdo será disponibilizado, de forma versionada, à medida em que for sendo produzido.

Quanto aos direitos de propriedade intelectual, foi realizada pesquisa de busca com os termos da pesquisa nas seguintes bases de propriedade intelectual: Instituto Nacional da Propriedade Industrial (INPI); LATIPAT-Espacenet; WIPO - Search International and Nacional Patent Collections; USPTO - United States Patent and Trademark Office. O resultado das buscas consta como anexo no projeto de pesquisa e não infringem direitos de propriedade intelectual.

3.2 Como serão gerenciadas as questões de direitos autorais e direitos de propriedade intelectual?

Orientação:

- **Direitos autorais** – Declarar quem será o proprietário dos direitos autorais de quaisquer dados que você coletar ou criar, junto com a(s) licença(s) para seu uso e reutilização. Considere também parceiros e as permissões para reutilizar dados de terceiros e quaisquer restrições necessárias ao compartilhamento de dados.
- **Perguntas orientadoras:**
  - ✓ Quem é o dono dos dados?
  - ✓ Como os dados serão licenciados para reutilização?
  - ✓ Haverá período de embargo ou restrições de acesso aos dados e por quê? (exemplo: para publicação ou solicitação de patente)

Preencher aqui:

## 4 Armazenamento e backup

### 4.1 Como os dados serão armazenados e preservados durante a pesquisa?

Orientação:

- **Armazenamento** – Relatar brevemente como os dados serão armazenados e os procedimentos de segurança, incluindo backup, durante a pesquisa. Indique com que frequência os dados serão copiados e em quais locais. Armazenar dados em discos rígidos de computador ou dispositivos de armazenamento externo sozinho é muito arriscado. Normalmente é melhor usar serviços de backup automático fornecidos por serviços de TI. Se optar por usar um serviço de terceiros, deve-se garantir que não entre em conflito com qualquer financiador, políticas institucionais ou departamentais como em termos de jurisdição legal em que os dados serão mantidos ou a proteção de dados confidenciais.
- **Perguntas orientadoras:**
  - Você tem armazenamento suficiente ou precisará incluir encargos para serviços adicionais?
  - ✓ Como será feito o backup dos dados?
  - ✓ Quem será responsável pelo backup e recuperação?

- ✓ Como os dados serão recuperados em caso de incidente?

Preencher aqui:

Uma vez que os dados da pesquisa serão coletados, tratados e criados a partir de um script Python usando um Caderno Jupyter, seu armazenamento e backup serão produzidos a cada ciclo de utilização, sendo registrados diretamente no repositório do GitHub. As entregas parciais serão então publicadas como um *release* no GitHub e espelhadas automaticamente no repositório do Zenodo, como já configurado.

Além disso, será também gerada uma cópia local no laptop de trabalho, atualizada a cada nova modificação na árvore de arquivos do repositório do GitHub, usando a ferramenta de sincronização GitHub Desktop.

A responsável pelo backup e recuperação dos dados será a pesquisadora. Em caso de incidentes, os dados podem ser recuperados na máquina local ou no repositório do GitHub, bem como novamente produzidos por meio do script Python adequado à reprodutibilidade da pesquisa.

#### 4.2 Como você gerenciará o acesso e a segurança?

Orientação:

- **Dados confidenciais** – Delinear todas as medidas de segurança adequadas e observar quaisquer padrões formais que você cumprirá (exemplo: ISO 27001) em caso de dados confidenciais (exemplos: dados pessoais que ainda não sejam de domínio público, informações confidenciais ou segredos comerciais).
- Perguntas orientadoras:
  - ✓ Quais são os riscos para a segurança dos dados e como eles serão gerenciados?
  - ✓ Como será controlado o acesso para manter os dados seguros?
  - ✓ Como você garantirá que os colaboradores possam acessar seus dados com segurança?
  - ✓ Se estiver criando ou coletando dados no campo, como você garantirá sua transferência segura para seus principais sistemas protegidos?

Preencher aqui:

Quanto ao gerenciamento do acesso e segurança, reitera-se que não serão utilizados dados confidenciais. Os riscos para a segurança dos dados serão gerenciados por meio de login e



senha na máquina de trabalho local e nos repositórios do GitHub e do Zenodo. Não haverá coleta de dados em campo e a pesquisa não conta com colaboradores, sendo de responsabilidade exclusiva da pesquisadora.

## 5 Seleção e Preservação

5.1 Quais dados tem valor a longo prazo e devem ser retidos, compartilhados e/ou preservados?

Orientação:

- **Preservação** – Como os dados para preservação serão selecionados e onde os dados serão preservados a longo prazo (exemplo: em um repositório de dados)? Lembre-se de considerar qualquer esforço adicional necessário para preparar os dados para compartilhamento e preservação, como alterar formatos de arquivo. Todos os dados coletados ou produzidos em projetos de pesquisa fomentados pelo CNPq serão depositados no repositório LattesData.
- **Perguntas orientadoras:**
  - ✓ Quais dados devem ser retidos para fins contratuais, legais ou regulamentares?
  - ✓ Como será decidido quais outros dados manter?
  - ✓ Quais são os usos de pesquisa previsíveis para os dados?
  - ✓ Por quanto tempo os dados serão retidos e preservados?

Preencher aqui:

Além de serem depositados no repositório LattesData, todos os dados criados serão preservados nos repositórios GitHub e Zenodo, conforme citado anteriormente, e serão mantidos enquanto estes repositórios estiverem disponíveis.

A previsibilidade de reuso dos dados criados se dá em diversas áreas do conhecimento, especialmente em pesquisas no campo das Ciências Humanas e Sociais que se dediquem a estudar o pensamento gramsciano no Brasil.

Os repositórios utilizados são mantidos por organizações comerciais (GitHub) e de pesquisa (Zenodo), sendo repositórios de livre acesso, sem custos para os usuários pesquisadores e, no caso do LattesData, os encargos são do CNPQ.

recursos e sistemas estarão disponíveis para permitir que haja curadoria dos dados de forma eficaz além do prazo de validade da concessão. Além disso, o dataset deverá ter os seus metadados também depositados no LattesData referenciando o depósito no repositório escolhido.

## 5.2 Qual é o plano de preservação de longo prazo para o conjunto de dados?

Orientação:

- **Plano de preservação** – Considere como os conjuntos de dados com valor de longo prazo serão preservados. Caso o repositório utilizado para preservação não seja o LattesData, o PGD deve demonstrar que os recursos e sistemas estarão disponíveis para permitir que haja curadoria dos dados de forma eficaz além do prazo de validade da concessão. Além disso, o dataset deverá ter os seus metadados também depositados no LattesData referenciando o depósito no repositório escolhido.
- **Perguntas orientadoras:**
  - ✓ Em qual repositório ou arquivo os dados, por exemplo, os dados serão armazenados?
  - ✓ Quais são os custos, se houver, com o repositório de dados ou arquivo selecionado?
  - ✓ Houve gasto de tempo e esforço para preparar os dados para compartilhamento/preservação?

Preencher aqui:

## 6 Compartilhamento de dados

### 6.1 Como você compartilhará os dados?

Orientações:

- **Métodos para compartilhar dados**– Como, quando e para quem os dados serão compartilhados? Se possível, mencione exemplos anteriores para mostrar um histórico de compartilhamento eficaz de dados. Os dados deverão ser compartilhados, no caso de beneficiários do CNPq em prazo não superior a 60 dias após o término do benefício.

- Perguntas orientadoras:

- ✓ Quando serão disponibilizados os dados?
- ✓ Se o registro de identificador DOI - Digital Object Identifier será feito para o conjunto de dados ou se haverá necessidade de atribuir individualmente identificadores arquivos, pois isto implica custos ao detentor do repositório LattesData.
- ✓ Caso o depósito não seja feito originalmente no LattesData, como será assegurado o registro de um identificador único e persistente (como um DOI - Digital Object Identifier) para cada conjunto de dados?

Preencher aqui:

Os dados serão disponibilizados ao longo da execução do projeto, à medida em que forem sendo trabalhados. O registro do DOI virá a cada *release* publicado no Zenodo, garantindo um identificador persistente para cada arquivo do projeto.

O Zenodo dispõe de um recurso de atribuição de DOI, com controle de versão, para cada arquivo ou conjunto de dados publicados.

Por exemplo, o presente Plano de Gestão de Dados, está publicado no Zenodo sendo o DOI da primeira versão do arquivo de trabalho o endereço:

- **Direito de Acesso e Licença** – Os dados coletados ou produzidos em projetos de pesquisa fomentados pelo CNPq e depositados no repositório LattesData serão de Acesso Aberto e Licença CC/BY ou CC0 ou CC BY-SA ou CC BY-NC, com padrão de citação conforme estabelecido no LattesData.

## 6.2 É necessária alguma restrição ao compartilhamento de dados?

Orientações:

- **Restrições previstas** – Descreva quaisquer dificuldades esperadas no compartilhamento de dados com valor reconhecido de longo prazo, juntamente com as causas e possíveis medidas para superá-las. Considere se um acordo de não divulgação forneceria proteção suficiente para dados confidenciais.
- Perguntas orientadoras:
  - ✓ Que ação será tomada para superar ou minimizar as restrições?
  - ✓ Por quanto tempo será preciso o uso exclusivo dos dados e por quê?

- ✓ Será necessário um acordo de compartilhamento de dados (ou equivalente)?

Preencher aqui:

## 7 Responsabilidades e Recursos

### 7.1 Quem será o responsável pela gestão dos dados?

Orientações:

- **Responsabilidade** – Quem será o responsável pela gestão dos dados de pesquisa? Antes do depósito, no caso de beneficiários do CNPq, o responsável pela gestão dos dados será o Coordenador do Projeto, e após o depósito no LattesData, caberá ao CNPq a gestão dos dados, conforme estabelecido no Termo de Depósito do LattesData.

Preencher aqui:

### 7.2 Quais são os detalhes do projeto e de autoria que precisam ser esclarecidos?

Orientações:

- **Projeto de pesquisa** – Definir e descrever aspectos da pesquisa. Elementos principais como o título, resumo e palavras-chave precisam ser fornecidos.
- **Perguntas orientadoras:**
  - ✓ Qual é o título do projeto?
  - ✓ Quais são as palavras-chave?
  - ✓ Qual é a instituição executora?
  - ✓ Qual é o início da vigência e o término da vigência?
  - ✓ Quais foram os recursos financeiros e suas fontes (quantidade, alocação, distribuição e organização de recursos em geral)?
  - ✓ Qual foi a chamada? Qual o número do processo?
  - ✓ Qual a área do conhecimento (baseado nas áreas CNPq) que sua pesquisa se enquadra?
  - ✓ Qual é a hipótese de pesquisa (objetivo dos dados)?

Preencher aqui:

- **Autoria** – Definir os dados pessoais essenciais sobre a autoria do conjunto de dados de pesquisa, tais como proponente, e-mail, identificador único de autor, instituição executora.
- **Perguntas orientadoras:**

- ✓ Qual o nome do responsável pelo conjunto de dados?
- ✓ Qual o endereço de email para contato sobre o conjunto de dados?
- ✓ Qual o identificador único do autor? (exemplo: ID Lattes, ORC ID)
- ✓ Qual o nome completo da instituição responsável pelos dados?
- ✓ Qual o telefone para contato sobre o conjunto de dados?

Preencher aqui:

**Importante: O proponente declara que deu ciência, a todos os envolvidos na geração ou coleta de dados, do compromisso de depósito dos dados de pesquisa no repositório LattesData e que todos estão de acordo.**