

**Diretrizes
para aplicação
dos princípios FAIR
à gestão e
reutilização de
dados**



Este guia propõe uma série de diretrizes para alinhar os esforços daqueles que produzem, gerem e usam dados no campo das ciências sociais e humanas, de modo a tornar os dados de investigação o mais reutilizáveis possível.

As orientações resultam do trabalho de mais de cinquenta profissionais envolvidos no projeto PARTHENOS, responsáveis por estudar pontos comuns na implementação de políticas e estratégias de gestão de dados de investigação. Foram analisados os resultados da pesquisa documental, bem como de questionários e entrevistas com especialistas selecionados e cerca de 100 documentos relacionados com políticas de gestão de dados (incluindo guias para formatos preferidos, políticas de revisão de dados e boas práticas adotadas tanto formal como tacitamente).

Com o foco na qualidade dos (meta)dados e repositórios, a equipa do PARTHENOS extraiu um conjunto de vinte diretrizes que são comuns às diferentes disciplinas.

Para facilitar a consulta, as diretrizes foram organizadas de acordo com os princípios de dados FAIR – Findable, Accessible, Interoperable e Reusable, - ou seja, dados Localizáveis, Acessíveis, Interoperáveis e Reutilizáveis, publicados pela primeira vez por FORCE11 (2016).

Cada uma das diretrizes do PARTHENOS inclui recomendações específicas dirigidas a dois grupos principais: por um lado, aos produtores e utilizadores de dados e, por outro lado, aos gestores dos arquivos de dados. Os símbolos gráficos indicam a qual dos destinatários as recomendações são endereçadas.

Depois de terminado o Projeto PARTHENOS, e a pedido de vários parceiros do projeto, as traduções destas diretrizes para outras línguas serão continuadas pelo projeto ARIADNEplus e pela ação SEADDA COST.



O símbolo lâmpada indica recomendações para produtores e utilizadores de dados, tais como investigadores e comunidades de investigação em História, Arqueologia, Estudos linguísticos e de Ciências Sociais.



O símbolo com as rodas dentadas indica recomendações para infraestruturas de investigação e arquivo de dados geridos por Institutos de Investigação e Instituições de Património Cultural.

20 ORIENTAÇÕES

para aplicação dos princípios FAIR à gestão e reutilização de dados



1

Invista em pessoas e em infraestruturas

Um pré-requisito importante para poder implementar as diretrizes deste guia, é investir em infraestruturas e na contratação e formação de especialistas em gestão de dados.



Conheça as melhores práticas de gestão de dados de investigação. Consulte os módulos de formação PARTHENOS ou consulte o Guia Especializado em Gestão de Dados do CESSDA.



Invista na contratação e na formação de especialistas em gestão de dados e defina um orçamento para investimentos em infraestruturas técnicas e em pessoal.

LOCALIZÁVEL

Os dados de investigação devem ser facilmente localizáveis tanto por humanos como por sistemas informáticos, com base em descrições obrigatórias dos metadados que permitam a descoberta de conjuntos de dados com interesse para a investigação.

2

Use identificadores persistentes

A localização dos dados é uma condição necessária para realizar outras etapas, desde o acesso à sua reutilização. Para ser localizável, qualquer objeto e conjunto de dados deve ser identificável de forma única e inalterável ao longo do tempo com um identificador persistente (PID). O PID continua a funcionar mesmo quando o endereço web de um recurso for alterado. Os PIDs podem ser de vários tipos, como por exemplo, Handle, DOI, PURL ou URN.



Cite o PID atribuído ao seu conjunto de dados nos resultados da sua investigação.



Selecione a forma apropriada do esquema de identificação persistente e atribua um PID a cada recurso. Use o Guia PID do NCDD para decidir qual o PID certo para sua infraestrutura de investigação.

3

Cite os dados de investigação

Se os dados da investigação tiverem um identificador persistente e forem citados de acordo com as normas da comunidade, os respetivos recursos ou conjuntos de dados serão localizáveis com mais facilidade.



Familiarize-se com as diretrizes para a citação de dados que são específicas do seu domínio de investigação ou disciplina e cite os dados em conformidade.



Forneça às comunidades de investigação informações sobre as melhores práticas relacionadas com a citação de dados e facilite a citação de dados pelos utilizadores, por exemplo, através de um botão com as palavras "Como citar este conjunto de dados".

4

Use identificadores de autor persistentes

Um identificador de autor persistente (por exemplo, VIAF, ISNI ou ORCID) é usado para criar ligações entre o autor e os conjuntos de dados, atividades de investigação e publicações, promovendo o seu reconhecimento e identificação.



Distinga-se de qualquer outro investigador ou grupo de investigação. Se ainda não o fez, solicite um identificador de autor e associe-o aos seus conjuntos de dados.



Mencione os identificadores do autor nos metadados.

5

Escolha um esquema de metadados apropriado

Os metadados são essenciais para tornar os dados localizáveis, especialmente os que são usados para citar e descrever dados. Um esquema de metadados é uma lista de elementos padronizados que fornece informações sobre um recurso, como por exemplo, o título, o identificador, o nome do autor ou uma data. Use esquemas de metadados existentes, que cumpram normas internacionais, para facilitar a troca de dados.



Para permitir a descoberta de conteúdos, descreva os dados de investigação da maneira mais consistente e completa possível. Inclua informações suficientes para que os dados possam ser acedidos e compreendidos ao longo do tempo. Se possível, use um esquema de metadados existente que se ajuste ao tipo de recurso ou conjunto de dados que pretende descrever.



Indique claramente qual o esquema de metadados que está a usar e partilhe-o com a comunidade de investigadores. Quando deposita os dados num repositório, considere a possibilidade de recolher metadados adicionais, como por exemplo, a proveniência dos dados, através de um formulário que deve acompanhar os dados.

ACESSÍVEL

Os dados da investigação devem ser facilmente acessíveis e recuperáveis com condições de acesso bem definidas, usando protocolos padrão de comunicação.

6

Escolha um repositório confiável

Um repositório certificado oferece um alojamento confiável para os seus conjuntos de dados. A certificação é uma garantia de que os dados são armazenados com segurança e permanecem disponíveis, localizáveis e acessíveis a longo prazo. Exemplos de certificação standards são CoreTrustSeal, Nestor Seal e certificação ISO 16363.



Torne seus dados acessíveis depositando-os num repositório confiável. Além disso, se usar as normas e padrões recomendados pelos repositórios (relacionados com os formatos de ficheiros, esquemas de metadados, etc.), poderá garantir que está a cumprir todos os requisitos para tornar os seus dados FAIR (Findable, Accessible, Interoperable, Reusable).



Indique claramente o nível de certificação do seu website. Se o seu repositório ainda não é certificado, indique como planeia garantir a disponibilidade, localização, acessibilidade e reutilização dos dados a longo prazo.

7

Indique de forma explícita a acessibilidade

As informações de acesso definem como um utilizador pode ter acesso a um conjunto de dados. Quando deposita dados num repositório, devem ser claros quais os tipos de acesso que pode escolher.



Ao escolher um tipo de acesso, considere os requisitos legais, políticas específicas da disciplina e protocolo de ética, quando aplicável. Sempre que possível opte por Acesso Aberto. Ao guardar dados pessoais, verifique se eles contêm alguma informação que possa levar à divulgação das identidades dos participantes, se estão de acordo com o que consentiram e que medidas tomou para proteger os dados. Se os seus dados não puderem ser publicados em Acesso Aberto, os respetivos metadados deverão ser, para permitir a descoberta dos dados.



Incentive a publicação de (meta)dados em Acesso Aberto. Opte por um acesso restrito apenas para (meta)dados confidenciais e que não devem ser publicamente acessíveis. Nesse caso, procure disponibilizar os (meta)dados através de procedimentos de acesso controlado e documentado.

8

Utilize o embargo de dados somente quando necessário

Durante o período de embargo dos dados, apenas a descrição do conjunto de dados é pública. Os dados não estão acessíveis e os metadados completos só ficarão disponíveis após um certo período de tempo.



Indique claramente porque motivo e por quanto tempo o embargo de dados é necessário. Disponibilize o acesso aos metadados o mais rapidamente possível.



Especifique quando é que um embargo de dados é permitido e que condições se aplicam.

9

Use protocolos de interoperabilidade padronizados

Ao usar protocolos de troca padronizados, as infraestruturas de investigação podem tornar os (meta)dados publicamente localizáveis e disponíveis para acesso, por exemplo, por motores de pesquisa, melhorando bastante a sua acessibilidade.



Use protocolos standard como SWORD, OAI-PMH, ResourceSync e SPARQL. Converta esquemas de (meta)dados em ficheiros XML ou RDF. Mantenha e publique um registo de protocolo de endpoints, indicando o caminho de acesso aos dados da investigação.

Para acelerar a descoberta e revelar novos conhecimentos, os dados de investigação devem poder ser facilmente combinados com outros conjuntos de dados, tanto por humanos como por sistemas informáticos.

INTEROPERÁVEL

10

Estabelecer APIs bem documentadas e executáveis por sistemas informáticos

Uma interface de programação da aplicação (API) bem documentada e executável por sistemas informáticos – através de um conjunto de procedimentos, protocolos e ferramentas para desenvolver software aplicativo – permite indexar, recuperar e combinar automaticamente (meta)dados de diferentes repositórios.



Documente bem as APIs e forneça o esquema do (meta)modelo de dados. Considere fornecer exemplos de como extrair com êxito dados de diferentes endpoints e como combiná-los com novos conjuntos de dados, que podem ser usados em novas investigações.

11

Use vocabulários bem definidos

A descrição dos elementos dos metadados deve seguir as diretrizes adotadas pelas comunidades científicas que usam vocabulários abertos, bem definidos e conhecidos. Esses vocabulários descrevem o significado exato dos conceitos e características que os dados apresentam.



Desde o início do seu projeto de investigação use vocabulários relevantes na sua área e enriqueça e estruture consistentemente os resultados da sua investigação.



Forneça à comunidade de investigação exemplos de vocabulários baseados nas especificidades do seu domínio de investigação.

12

Documente o modelo de metadados

Documentar de forma clara os modelos de metadados ajuda os programadores a comparar e mapear diferentes esquemas de metadados.



Publique os modelos de metadados usados na sua infraestrutura de investigação. Documente as especificações técnicas e defina classes (grupos de itens que possuem propriedades comuns) e propriedades (elementos que expressam os atributos duma seção de metadados, bem como as relações entre diferentes componentes dos metadados). Liste as propriedades obrigatórias e recomendadas para o mapeamento dos metadados.

13

Estabeleça e use normas de dados interoperáveis

O uso de normas adotadas por uma comunidade grande aumenta a possibilidade de partilha, reutilização e combinação de conjuntos de dados.



Verifique quais as normas e padrões usados no repositório onde pretende depositar os seus dados. Estruture o seu conjunto de dados de acordo com esse formato, desde o início do seu projeto de investigação.



Indique claramente quais os standards que são usados pela sua instituição, partilhe-os com a comunidade e mantenha-os especialmente numa perspetiva da interoperabilidade. Alguns bons exemplos são CMDI, SIKB0102, METS e CIDOC-CRM.

14

Estabeleça procedimentos para melhorar a qualidade dos dados

Para melhorar a qualidade dos (meta)dados e, conseqüentemente, a sua interoperabilidade, estabeleça procedimentos automáticos para limpar, depositar e enriquecer os (meta)dados.



Estabeleça procedimentos para minimizar o risco de erros na recolha de dados. Por exemplo, selecione uma data de um calendário em vez de digitá-la manualmente.



Invista em ferramentas para ajudar a limpar (meta)dados e converter dados em formatos normalizados e interoperáveis. Desenvolva fluxos de trabalho e soluções de software para executar procedimentos automáticos, por exemplo, usando ferramentas de aprendizagem automática.

15

Recomende e use formatos de arquivo sustentáveis a longo prazo

Todos os ficheiros de dados armazenados num repositório devem estar num formato internacional aberto e normalizado, para garantir a interoperabilidade a longo prazo, em termos de acessibilidade, reutilização e sustentabilidade dos dados.



Pense em usar formatos de ficheiros sustentáveis, desde o início do seu projeto de investigação. Siga as recomendações sugeridas pelo repositório de dados, usando os formatos preferidos, independentes do software, programadores ou fornecedores específicos.



Incentive o uso de formatos considerados adequados para preservação a longo prazo, como ficheiros PDF-A, CSV e MID / MIF. Forneça uma visão detalhada e fácil de identificar para os formatos aceites.

Os dados de investigação devem estar preparados para pesquisas e processamentos futuros, realçando que as descobertas podem ser replicadas e que novas investigações podem ter por base resultados anteriormente obtidos.

REUTILIZÁVEL

16

Documente sistematicamente os dados

Os dados devem ser sistematicamente documentados para tornar claro o que se pode esperar de um conjunto de dados ou repositório. A transparência sobre o que é ou não encontrado nos dados confere confiabilidade e, conseqüentemente, também promove a reutilização.



Forneça manuais que incluam a metodologia usada, a lista de abreviaturas, a descrição das lacunas nos dados, a configuração da base de dados, etc.

17

Respeite as convenções de terminologia

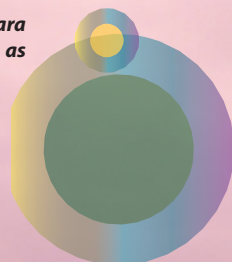
Use uma convenção terminológica precisa e consistente - um esquema geralmente aceite para nomear ficheiros de dados - facilita significativamente que gerações futuras de investigadores possam recuperar, compreender e aceder a recursos e conjuntos de dados.



Consulte as políticas e práticas recomendadas para a sua disciplina ou domínio de investigação para identificar a convenção terminológica mais adequada.



Indique claramente quais são as práticas recomendadas para criar e aplicar convenções específicas relacionadas com as terminologias de ficheiros.



18

Use formatos de ficheiros comuns

A reutilização de dados aumenta quando são usados os formatos padrão mais populares na sua comunidade.



Para partilhar os seus dados, publique os formatos atuais mais populares juntamente com os formatos de arquivo, como por exemplo, formatos Excel (xlsx) e CSV ou ESRI ao lado de ficheiros MID/MIF.



Publique os dados nos formatos comuns e nos formatos de arquivo recomendados, caso sejam diferentes.

19

Mantenha a integridade dos dados

Os dados recolhidos na investigação devem ser idênticos aos dados acedidos posteriormente. Para garantir a autenticidade dos dados devem ser realizadas verificações de integridade.



Implemente um método para controle de versão. Para garantir a autenticidade de cada conjunto de dados é da maior importância que todas as alterações à versão original sejam devidamente documentadas.



Para identificar se um arquivo foi modificado, é essencial registar a proveniência - a origem dos dados e as alterações feitas ao longo do tempo - e comparar qualquer cópia com o original. A integridade dos dados tanto pode ser verificada por meio de uma impressão digital, como por um código de controlo do tipo de soma de verificação, ou por comparação direta entre dois ficheiros. Forneça um mecanismo para destacar as diferentes versões, por exemplo, adicionando a versão do identificador como um parâmetro de pesquisa.

20

Licença para reutilização

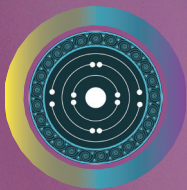
Para permitir a maior reutilização possível, deve ser claramente referido quem possui os direitos sobre os (meta)dados e que tipo de licença se aplica.



Antes de publicar seus dados de investigação, saiba quem possui os direitos dos (meta)dados.



Comunique de forma clara e transparente as licenças para reutilização dos (meta)dados e num formato legível por sistemas informáticos. Para melhorar a interoperabilidade, tente mapear as suas licenças com estruturas que já são amplamente adotadas, como o Creative Commons.



 **ARIADNE**_{plus}

 **SEADDA**
Saving European Archaeology from the Digital Dark Age



PARTHENOS e ARIADNEplus são projetos Horizonte 2020 financiados pela Comissão Europeia. SEADDA é uma AÇÃO COST (CA18128) financiada pelo Programa-Quadro Horizonte 2020 da União Europeia. As visualizações e opiniões expressas nesta publicação são da responsabilidade exclusiva do autor e não refletem necessariamente as opiniões da Comissão Europeia.

Guia (versão original de dezembro de 2018, tradução de agosto de 2020) licenciado sob a licença Creative Commons CC BY 4.0. Design: Verbeeldingskr8. <https://doi.org/10.5281/zenodo.3937183>