

Fórum de Gestão de Dados de Investigação

Planos de Gestão de Dados na Prática

Workshop – 9° Fórum GDI – 25 novembro 2022

Grupo de Trabalho
Formação e competências
para Gestão e Dados FAIR



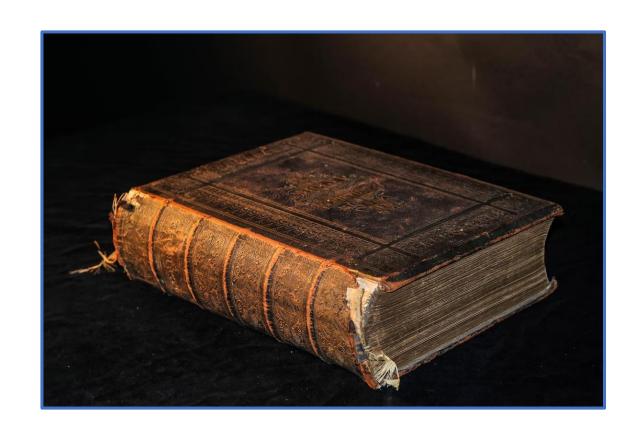
Programa

- Introdução
 - Plano de Gestão de Dados (PGD)
 - Argos ferramenta para a criação de PGD
 - Apresentação do desafio prático
- Exercício prático de criação de um Plano de Gestão de Dados com a utilização do Argos

O conceito de **Plano de Gestão de Dados** (**PGD**) tem origem no final dos anos 60 do século XX. No entanto só no século XXI é que viria a ser adotado pela comunidade científica na generalidade.

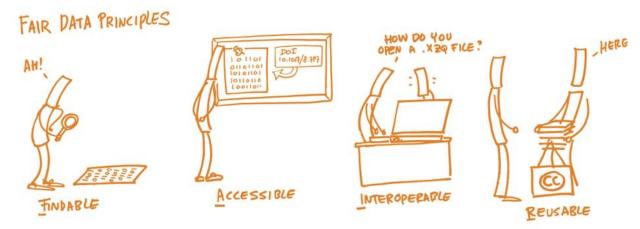
No século XXI, torna-se óbvio para a comunidade científica que as políticas das agências de financiamento não se alinham com os desafios criados pela revolução digital.

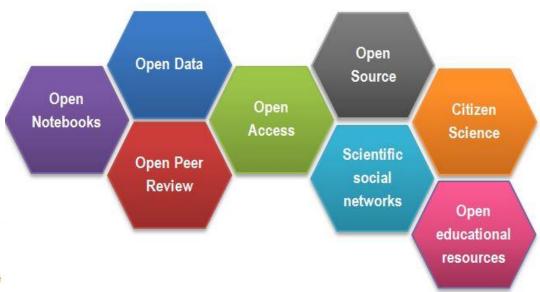
- Crescimento exponencial na produção de dados
- Financiamento para repositórios digitais
- Diretrizes para a preservação de longo termo
- Formatos não standardizados



É reconhecida pela comunidade científica a importância de fornecer acesso aberto aos dados de investigação, em particular aqueles resultantes de projetos financiados por fundos públicos.

Ganha força movimento em favor da ciência aberta e da adoção dos **princípios FAIR** por parte da comunidade científica.





https://www.fosteropenscience.eu/content/what-open-science-introduction

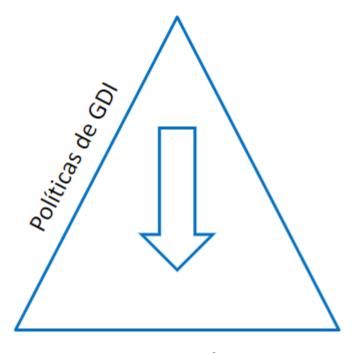
É consensual que quaisquer políticas de gestão de dados de investigação terão que ser aplicadas através de uma abordagem top down.

Sendo que o **conceito de PGD** surge como uma das **ferramentas** para a implementação desse tipo de abordagem.

Em particular com a ideia de encorajar a submissão um PGD juntamente com candidaturas a financiamento.

Esta tendência tem continuado ao longo da última década, com cada vez mais agências de financiamento a adotar essa prática.

Agências de Financiamento

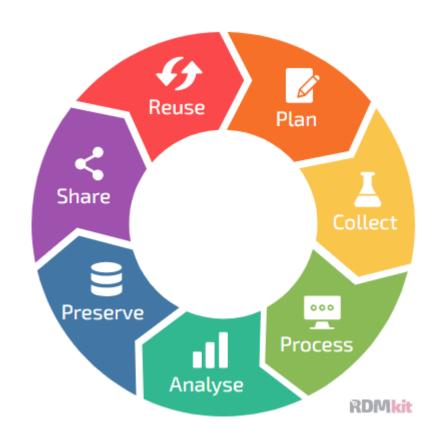


Investigadores

Um PGD é um documento formal que tem como objetivo fomentar o planeamento e servir de referência durante a execução de atividades de gestão de dados.

Em particular antecipando as **necessidades e requisitos ao longo de um projeto (de investigação)** numa determinada equipa, instituição ou organização.

Um PGD deve descrever políticas e métodos que são aplicados aos dados durante as múltiplas etapas do ciclo de vida dos dados (de investigação).



https://rdmkit.elixir-europe.org/data_life_cycle

O PGD está para a gestão de dados de investigação como as fichas técnicas estão para a restauração.

Fichas técnicas => Descrição detalhada de como preparar receitas

PGD => Descrição detalhada de como gerir dados de investigação

TARTAR SAUCE RECIPE TYPE: Batch 5 gallons SHELF LIFE: 5 days CONTAINER: 20-qt. Lexan TOOLS: Commercial mixer S/S mixing bowl Wire whip attachment POSITION(S) Prep cook

BLUE FISH GRILL RECIPE BOOK

Gallons Gallons	Mayonnaise
Gallons	
Cumons	Sour cream
Cup	Dijon mustard
Cups	Shallots (fine-chopped)
Cup	Dill relish
Cups	Fresh dill weed (fine-chopped)
Cups	Fresh parsley (fine-chopped)
TBSN	Tarragon leaves
TBSN	Hot sauce
Cup	Salt
TBSN	Black pepper
9	
18	
	Cup Cups Cups TBSN TBSN



2.	Using the wire whip attachment, mix until fluffy and well-blended.
3.	Cover and apply date label.
4.	Refrigerate.

1. Combine all ingredients into mixing bowl for

commercial mixer.

A cenoura:

Os PGD são documentos que **agregam** em **si toda a informação relativa à política de GDI** de um determinado projeto (de investigação).

São artefactos são o **produto** e simultaneamente **estimulam bom planeamento.**

Permitem a **rastreabilidade** entre atividades e atores.

Promovem o cumprimento dos princípios FAIR.

Potenciam a **automação** de atividades de gestão de dados.

Podem **acelerar** e **simplificar** os processos de investigação.



O pau:

Muitas agências de financiamento **encorajam a submissão de PGDs** juntamente com candidaturas a financiamento.

Esta política de **encorajamento** está em transição para **requerimento**.

Com os PGD as agências de financiamento procuram compreender como irão ser cumpridos os princípios FAIR.

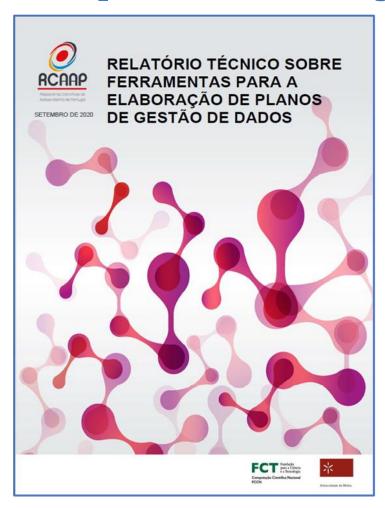
Os **PGD** devem que seguir templates definidos pelas agências de financiamento.

Permitem a **monitorização** e **validação** das políticas descritas, por parte das agências de financiamento.



Argos - Ferramenta para a criação de PGD

Disponibilização de um Sistema de PGDs



https://doi.org/10.21814/1822.67530

Objetivo

1

Avaliação das ferramentas disponíveis para a elaboração de PGDs



Serviço de âmbito nacional

Ferramentas para a elaboração de PGDs















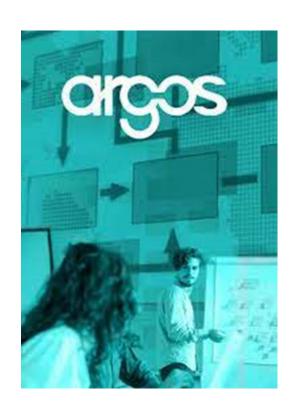






- DMPonline
 - https://dmponline.dcc.ac.uk/
- ARGOS
 - https://argos.openaire.eu/
- Data Stewardship Wizard https://ds-wizard.org/
- DMPTool
 - https://dmptool.org/
- OpenDMP
 - https://gitlab.eudat.eu/dmp/OpenAIRE-EUDAT-DMP-service-pilot
- ezDMP
 - https://ezdmp.org
- RDMO RDM Organiser
 - https://rdmorganiser.github.io/en/
- easyDMP
 - https://easydmp.eudat.eu
- DMP ARIADNEplus
 - https://vast-lab.org/dmp/ariadneplus/form/

Solução





Serviço Cloud

Argos

Plan and follow your data

Create machine actionable DMPs.

Configure to best fit your dicsipline.

Link to EOSC components out of the box.

Share easily in your repository.

Bring your Data Management Plans closer to where data are generated, analysed and stored.



https://argos.openaire.eu/

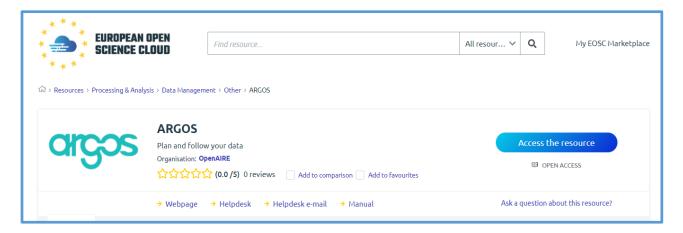




- Colaboração entre Infraestruturas Europeias ->
- Projetos -> Via OpenAIRE
- Flexível -> Requisitos e necessidades do financiador
- Simples -> Focada na interação com o utilizador

Argos - Cont.

- Adaptável -> Diferentes disciplinas e domínios
- maDMPs -> Criação de planos acionáveis por máquina
- Versão em Português
- Ligada à European Open Science Cloud (EOSC) -> Catálogo de serviços:



Estrutura

Plano de Gestão de Dados

- Pode ser exportado em vários formatos
- Composto por Datasets

- Nome, restrito ou público, instituição, investigador, financiador, grant, licença
- Seleção do Dataset/Modelo de dados

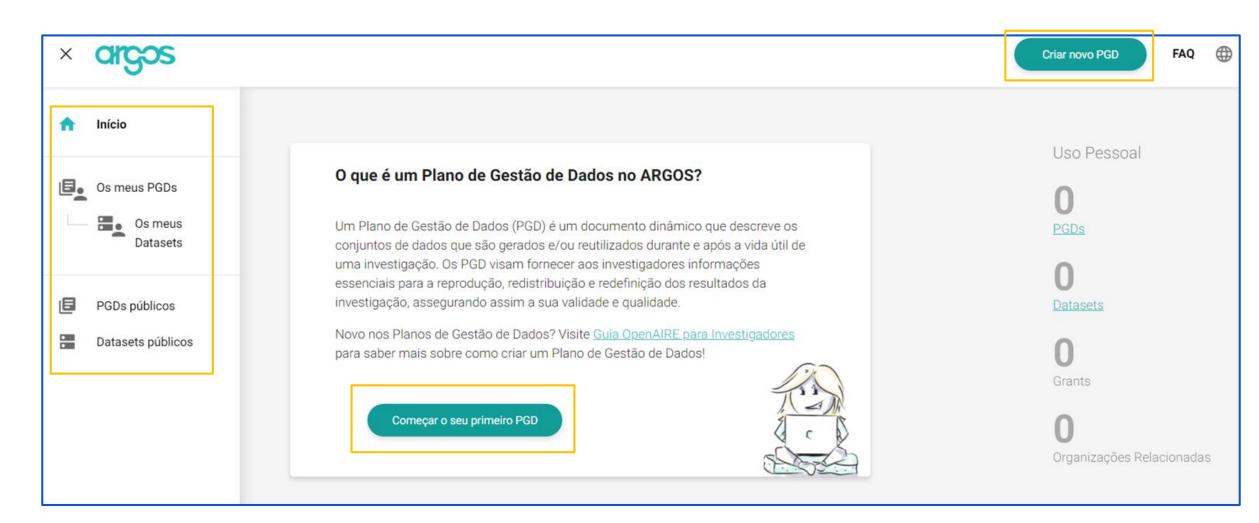
Dataset

 Conjunto de questões prédefinidas de acordo com modelo de instituição e/ou financiador



 Como serão tratados, armazenados e validados os dados gerados ou reutilizados durante o projeto de investigação

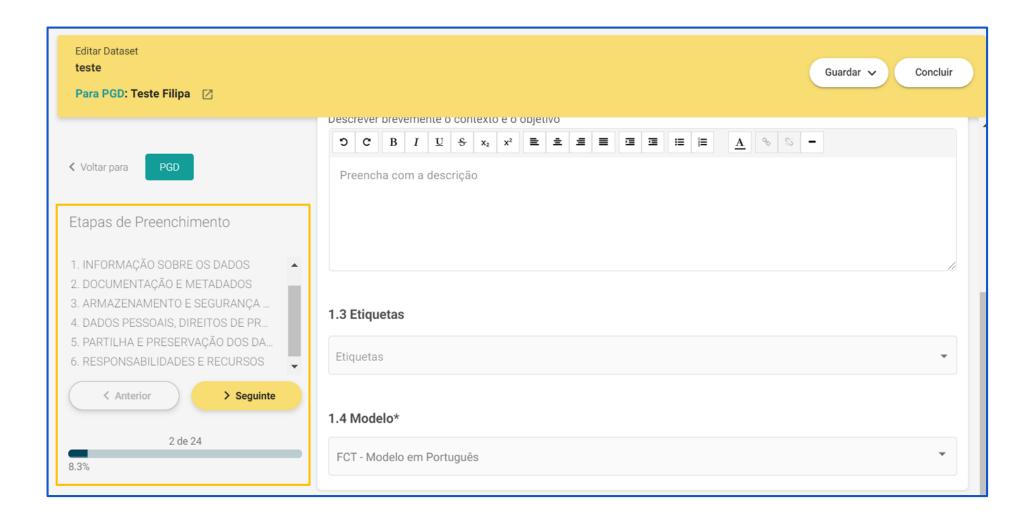
Interface



PGD



Dataset



Funcionalidades importantes

→ Login/autenticação simples:

Entrar na conta

D EDAT

CG P M

ZEMOD

Não é necessário criar uma conta!

Funcionalidades importantes - Cont.

→ Redação colaborativa:



→ Suporta exportações em formato JSON:

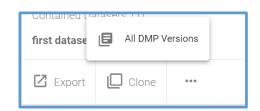


```
"dmp": {
    "contact": {
       "mbox": "elli.libd@gmail.com",
        "name": "Elli Papadopoulou",
            "identifier": "a7c440e8-4e29-4d84-812a-1b2d5cb605f5".
            "type": "other"
    "created": "18-09-2019",
    "dataset": [
            "data_quality_assurance": [
                "identifier": "56ac2eba-ea09-49cf-bba3-87fe48a10827",
                "type": "other"
            "issued": "9/18/19",
            "keyword": [
            "language": "en",
            "metadata": [
            "personal_data": "unknown".
            "security_and_privacy": [
            "sensitive_data": "unknown",
            "technical_resource": [
            "title": "Horizon 2020 Dataset Description"
```

Funcionalidades importantes - Cont.

→ Fornece DOIs e versionamento dos PGDs:





→ Adota o RDA DMP Common Standard:



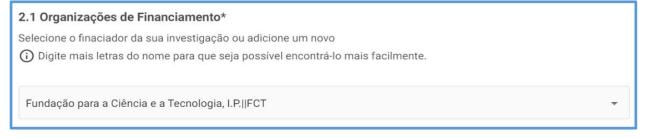
→ Usufrui das APIs do OpenAIRE, facilitando o preenchimento dos PGDs :

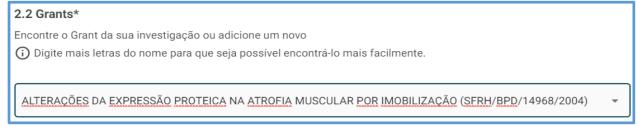


APIs - Fontes externas

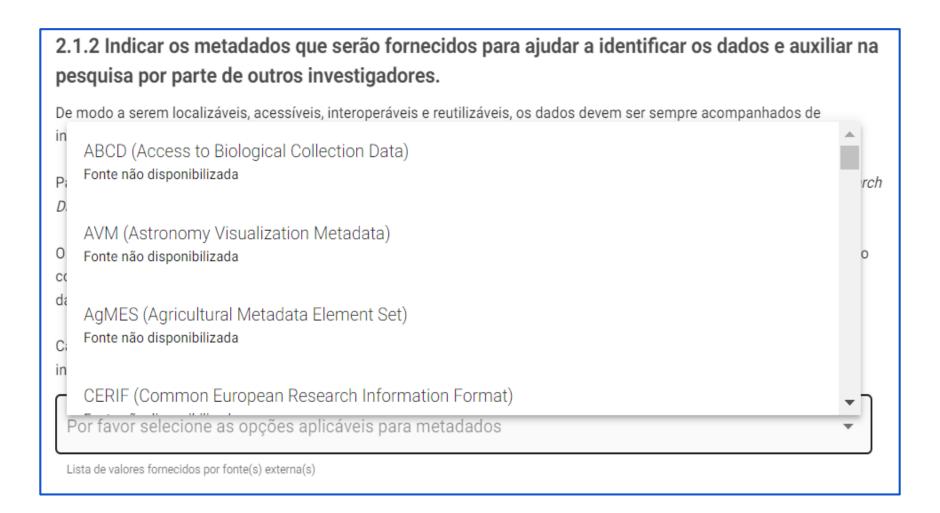
1.5 Investigadores Indique os nomes das pessoas que geraram, processaram e analisaram os dados descritos no PGD. (i) Digite mais letras do nome para que seja possível encontrá-lo mais facilmente. Selecionar investigadores Filipa Pereira (orcid:0000-0002-5732-9996)

1.6 Organizações Adicione aqui os nomes das organizações que contribuem para a criação e revisão dos PGDs i Digite mais letras do nome para que seja possível encontrá-lo mais facilmente. Selecione a organização UNIVERSIDADE DO MINHO

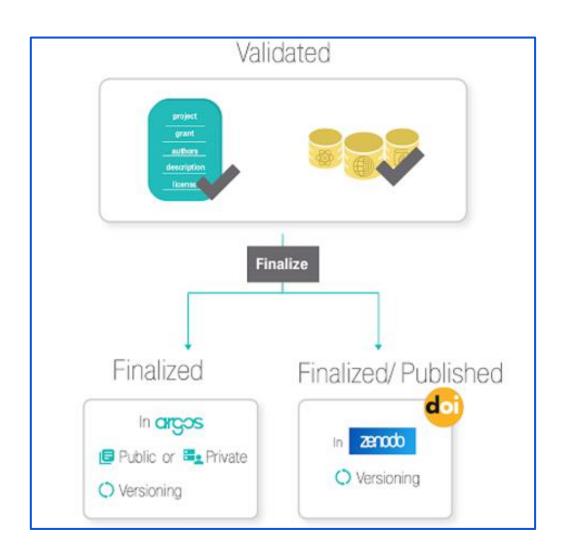




APIs - Fontes externas



Finalização



→ Após validação e submissão:

Argos - modo privado ou público

Ou

Zenodo - publicação

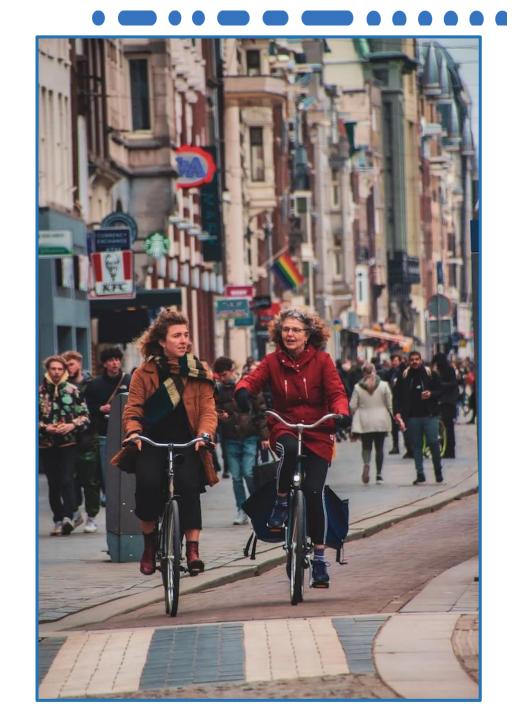
Exercício prático de criação de um PGD com a utilização do Argos

(Inspirado no projeto UCL 100 cyclist study https://www.ucl.ac.uk/civil-environmental-geomatic-engineering/research-projects/2021/jun/100-cyclists-project)

O Projeto Quase procura compreender os **fatores de risco associados a desastres de bicicleta**, no curso do seu uso em deslocações entre casa e os locais de trabalho.

Foram selecionados **100 voluntários** que terão montado na sua bicicleta um **dispositivo** que recolhe **dados de vídeo** e de **informação GPS**.

Na situação de o voluntário sofrer um 'quase acidente' é pedido que assinalem o evento pressionando um botão no dispositivo, que procederá ao registo permanente dos dados referentes aos 5 minutos que antecederam o evento.

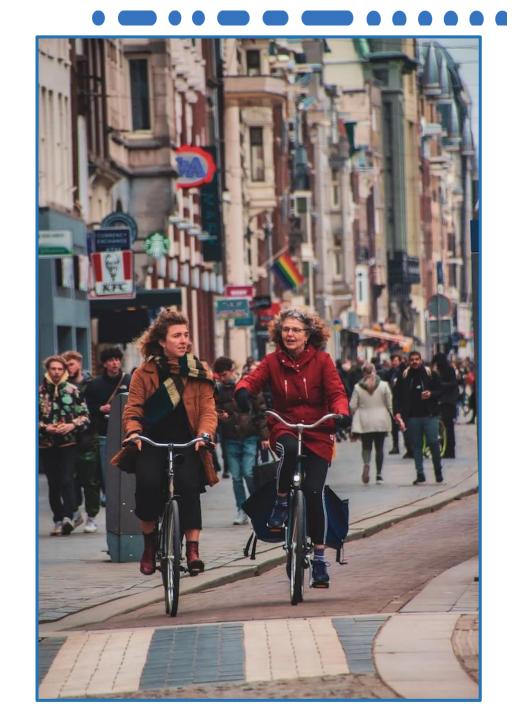


A recolha de dados terá início no mês de Janeiro de **2023** e final no mês de **Janeiro de 2025**.

O dispositivo tem capacidade estimada para armazenar os dados referentes a 5 eventos, sendo feita uma sincronização automática com o servidor de dados da instituição durante o carregamento do mesmo.

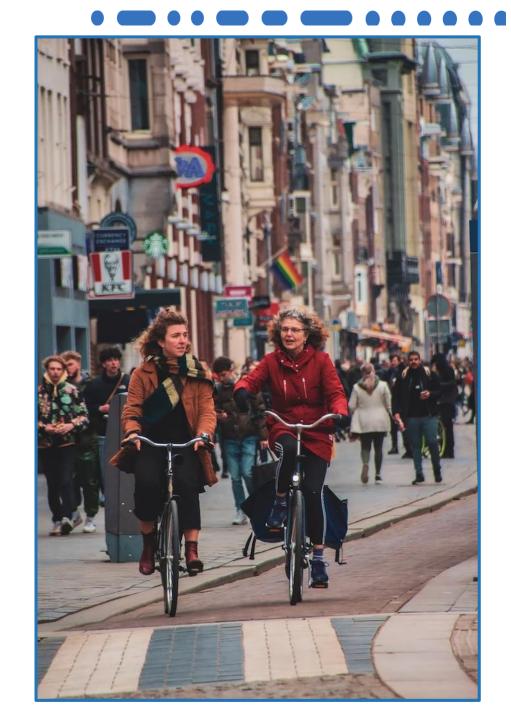
Ao serem depositados no servidor de dados da instituição, todos os dados são acompanhados de um manual de operações que define as condições de recolha dos dados.

O servidor de dados da instituição opera com uma infraestrutura redundante, sendo que os dados são replicados em duas instâncias do servidor de dados da instituição.



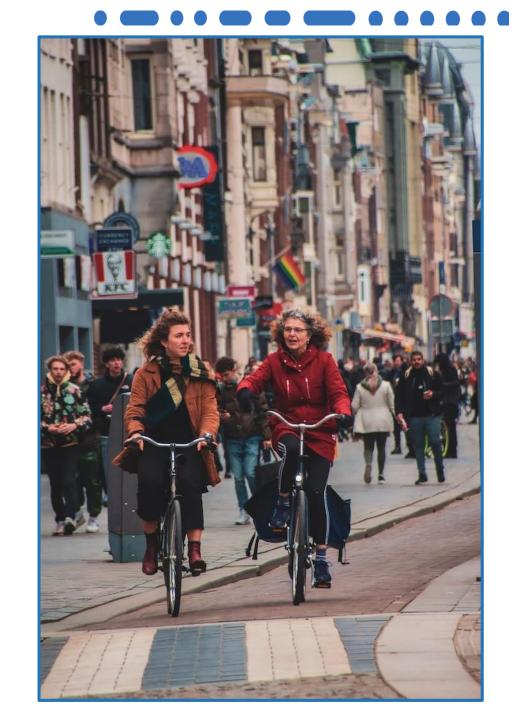
O Projeto Quase é um trabalho de investigação liderado pela **Universidade de Évora**, sendo **financiado** pela FCT através do grant FCT/2023/12345, com duração prevista de 2 anos.

Os dados gerados no âmbito do Projeto Quase são propriedade da Universidade de Évora. Sendo que estarão disponíveis para consulta mediante acesso controlado.



O **investigador principal** é a investigadora convidada Filomena Borba. Também na equipa estão presentes:

- André Vieira, João Cardoso e Filipa Pereira com tarefas de gestão de dados;
- Patrícia Miranda responsável pelo armazenamento durante a execução do projeto;
- Salima Rehemtula com tarefas de Data Protection Officer;
- Patrícia Dias de Melo responsável pela administração técnica do projeto.



Serão considerados os seguintes conjuntos de dados:

Dados Espaciais

Imagens com dados de georreferencia ção de origem externa

Formatos:

- .tif
- .tfw

Volume:

~1TB

Vídeo

Imagens vídeo do dispositivo (resolução 4k [UHD])

Formatos:

.mp4

Volume:

~15GB

Dados de GPS

Imagens com dados de gps do dispositivo

Formatos:

.gpx

Volume:

• < 1GB

Dados de **Eventos**

Ficheiro que associa evento e ficheiros de vídeo e dados GPS.

Formatos:

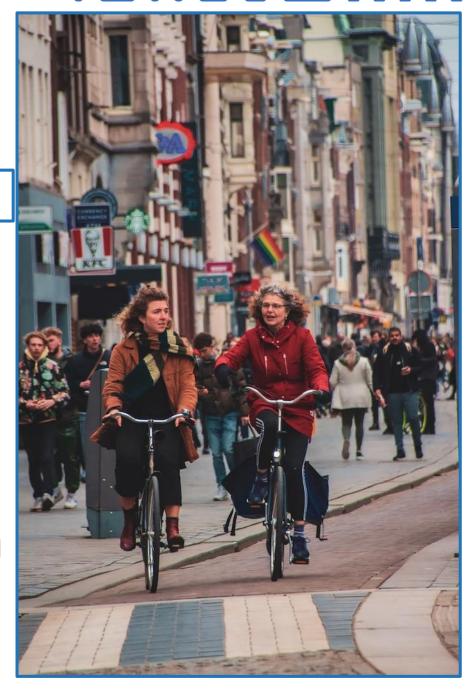
.txt

Volume:

• < 1GB

Metadados:

- Marca temporal
- Identificad or
- Ficheiro de vídeo
- Ficheiro de gps



Serão considerados os seguintes conjuntos de dados:

Identificação de Ciclistas

Tabela de identificação dos ciclistas

Formatos:

.CSV

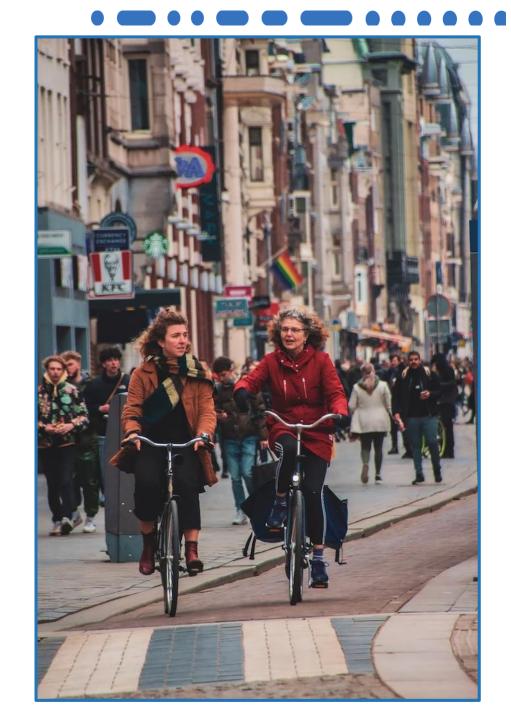
Volume:

• < 1GB

Metadados:

- Nome do voluntário (String) Idade do voluntário

- (String)
 Número de cartão de cidadão (String)
 Identificador (ID atribuído ao voluntário)



Argos



https://argos.openaire.eu/exploreplans/publicOverview/917f6b22-81b1-498f-9700-5cc801bd46fc



https://videoconf-colibri.zoom.us/j/5753471167

Tópicos do PGD da FCT

- Informação sobre os Dados
- Documentação e Metadados
 - Armazenamento e Segurança
 - Dados Pessoais e Direitos de Propriedade Intelectual
 - Partilha e Preservação a Longo Prazo
 - Responsabilidades e Recursos

Informação sobre os Dados

Qual a origem dos Dados?

Em geral os dados podem ter origem em

- Dados existentes
- Reutilização de dados

```
mestamp": "2017-06-03718:42:18
schars": "5022", "message": "n.Requesting the standard of the s
```

Informação sobre os Dados



Formato dos Dados

- Que tipo de extensão de ficheiro se pode usar para armazenar os dados?
- Porque utilizamos um determinado formato? a experiência do GAI
- normas dos repositórios
 Alguns repositórios fornecem listas dos formatos preferenciais
 - Preferir formatos não proprietários
- Facilitam a partilha e a preservação a longo prazo



Fonte: https://go.nature.com/2DELeN6



Reutilização dos Dados

Como proceder?

- Que dados vou reutilizar?
- Quais as condições para a sua reutilização?
- Existem restrições ao seu uso?
- De onde vêm os dados?



Vão ser reutilizados dados no projeto?

Resposta:

- ☐ Sim
- □ Não

Se sim:

- ☐ Que dados serão reutilizados e sob que condições de utilização?
- ☐ Descrever o tipo de dados que serão reutilizados e o seu formato

Serão considerados os seguintes conjuntos de dados:

Dados Espaciais

Imagens com dados de georreferencia ção de origem externa

Formatos:

- .tif
- .tfw

Volume:

• ~1TB

Vídeo

Imagens vídeo do dispositivo (resolução 4k [UHD])

Formatos:

• .mp4

Volume:

• ~15GB

Dados de GPS

Imagens com dados de gps do dispositivo

Formatos:

• .gpx

Volume:

• < 1GB

Dados de Eventos

Ficheiro que associa evento e ficheiros de vídeo e dados GPS.

Formatos:

• .txt

Volume:

• < 1GB

Metadados:

- Marca temporal
- Identificad or
- Ficheiro de vídeo
- Ficheiro de gps



O que ter em conta na concepção dos Dados?

- Metodologias de software
- Descrição dos dados
- Referir se foi considerada a utilização dos dados

Se não foi, qual o motivo?



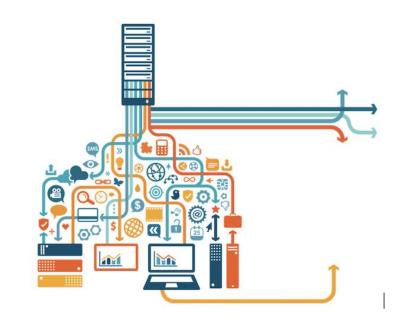
Vão ser gerados novos dados no projeto?

Resposta:

- ☐ Sim
- □ Não

Se sim:

- Descreva o tipo de dados que espera que a sua investigação venha a gerar.
- Descreva o formato de dados que espera que a sua investigação venha a gerar.
- Que metodologia ou software serão utilizados, caso sejam produzidos novos dados?



Serão considerados os seguintes conjuntos de dados:

Dados Espaciais

Imagens com dados de georreferencia ção de origem externa

Formatos:

- .tif
- .tfw

Volume: • ~1TB Vídeo

Dados de GPS

Dados de Eventos

Imagens vídeo do dispositivo (resolução 4k [UHD])

Formatos:

.mp4

Volume:

• ~15GB

Imagens com dados de gps do dispositivo

Formatos:

• .gpx

Volume:

• < 1GB

Ficheiro que associa evento e ficheiros de vídeo e dados GPS.

Formatos:

• .txt

Volume:

• < 1GB

Metadados:

- Marca temporal
- Identificad or
- Ficheiro de vídeo
- Ficheiro de gps



Serão considerados os seguintes conjuntos de dados:

Identificação de Ciclistas

Tabela de identificação dos ciclistas

Formatos:

.CSV

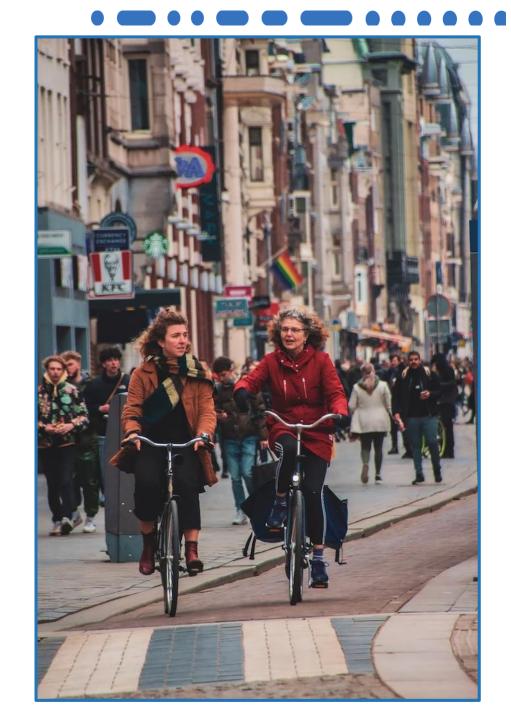
Volume:

• < 1GB

Metadados:

- Nome do voluntário (String) Idade do voluntário

- (String)
 Número de cartão de cidadão (String)
 Identificador (ID atribuído ao voluntário)



Volume dos Dados

Informação acerca da quantidade de dados que serão depositados. Qual é a capacidade de armazenamento que estima vir a necessitar para os guardar os dados, no total?

Resposta

0 - 10 GB

10 – 100 GB

100 – 1000 GB

> 1000 GB



Serão considerados os seguintes conjuntos de dados:

Dados Espaciais

Imagens com dados de georreferencia ção de origem externa

Formatos:

- .tif
- .tfw

Volume:

• ~1TB

Vídeo

Imagens vídeo do dispositivo (resolução 4k [UHD])

Formatos:

Volume:

.mp4

~15GB

Dados de GPS

Imagens com dados de gps do dispositivo

Formatos:

• .gpx

Volume:

• < 1GB

Dados de Eventos

Ficheiro que associa evento e ficheiros de vídeo e dados GPS.

Formatos:

• .txt

Volume:

• < 1GB

Metadados:

Marca temporai

- Identificad or
- Ficheiro de vídeo
- Ficheiro de gps



Serão considerados os seguintes conjuntos de dados:

Identificação de Ciclistas

Tabela de identificação dos ciclistas

Formatos:

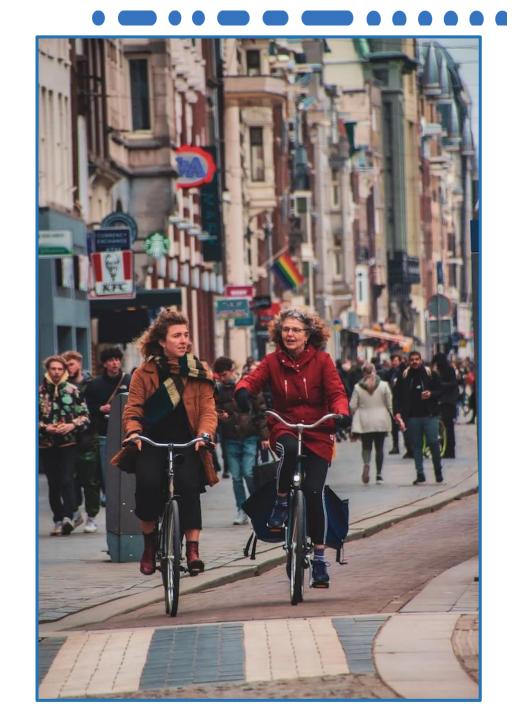
.csv

Volume:

• < 1GB

Metadados:

- Nome do voluntário (String) Idade do voluntário
- (String)
 Número de cartão de cidadão (String)
 Identificador (ID atribuído ao
- voluntário)



Tópicos do PGD da FCT

- Informação sobre os Dados
- Documentação e Metadados
 - Armazenamento e Segurança
 - Dados Pessoais e Direitos de Propriedade Intelectual
 - Partilha e Preservação a Longo Prazo
 - Responsabilidades e Recursos

- Que metadados e documentação acompanharão os dados?
 (ex.: a metodologia de recolha de dados e o modo de organização dados)
- Que medidas de controlo da qualidade dos dados serão utilizadas?
 - (para assegurar a correta gestão dos dados)

Documentação

Que documentação será necessária para permitir a reutilização dos dados. Informação sobre:

- metodologia utilizada para recolher os dados,
- informação analítica e processual,
- definições de variáveis,
- unidades de medida, etc.

Como é que a informação será capturada e onde será registada. Exemplos:

base de dados, ficheiro de texto 'readme', cadernos de laboratório, etc.

Como é que os dados serão organizados durante o projeto. Exemplo:

convenções adotadas, controlo de versões e estrutura de pastas.

Indicar que documentação irá acompanhar os dados.

Utilização de manual de operações que define as condições de recolha dos dados.

Projeto Quase

A recolha de dados terá início no mês de Janeiro de 2023 e final no mês de Janeiro de 2025.

O dispositivo tem capacidade estimada para armazenar os dados referentes a 5 eventos, sendo feita uma sincronização automática com o servidor de dados da instituição durante o carregamento do mesmo.

Ao serem depositados no servidor de dados da instituição, todos os dados são acompanhados de um manual de operações que define as condições de recolha dos dados.

O servidor de dados da instituição opera com uma infraestrutura redundante, sendo que os dados são replicados em duas instâncias do servidor de dados da instituição.

Metadados

De modo a serem FAIR (Localizáveis, Acessíveis, Interoperáveis e Reutilizáveis), **os dados devem ser sempre acompanhados de informação descritiva** sob a forma de metadados.

Aconselha-se a **utilização de standards de metadados**, como os que a RDA disponibiliza no seu Diretório: *Metadata Standards Catalog* (https://rdamsc.bath.ac.uk/)

O depósito de dados num repositório certificado / confiável envolve o fornecimento de informação sobre os dados de acordo com um **esquema de de metadados**.

Existem metadados associados aos dados?

⊠Sim □Não

Indicar os metadados que serão fornecidos para ajudar a identificar os dados e auxiliar na pesquisa por parte de outros investigadores.

São considerados os seguintes campos:

- Marca temporal (DateTime)
- Identificador (ID atribuído ao voluntário)
- Ficheiro de vídeo (mp4)
- Ficheiro de gps (gpx)

Serão considerados os seguintes conjuntos de dados:

Dados Espaciais

Imagens com dados de georreferencia ção de origem externa

Formatos:

- .tif
- .tfw

Volume:
• ~1TB

Vídeo

Imagens vídeo do dispositivo (resolução 4k [UHD])

Formatos:

• .mp4

Volume:

• ~15GB

Dados de GPS

Imagens com dados de gps do dispositivo

Formatos:

• .gpx

Volume:

• < 1GB

Dados de Eventos

Ficheiro que associa evento e ficheiros de vídeo e dados GPS.

Formatos:

• .txt

Volume:

• < 1GB

Metadados:

- Marca temporal
- Identificad or
- Ficheiro de vídeo
- Ficheiro de gps



Por favor identifique o(s) standard(s) de metadados que se aplica(m).

(Aconselha-se a utilização de standards de metadados, como os que a RDA disponibiliza no seu Diretório. Sabendo qual o repositório onde irá depositar os seus dados, poderá indicar aqui os metadados que são disponibilizados e adequados. Por exemplo: DataCite, DDI)

Por favor identifique o(s) vocabulário(s) de metadados que se aplica(m).

(Selecionar uma das opções ou adicionar manualmente. Deverá identificar os vocabulários de metadados que se aplicam ao seu conjunto de dados)

Por favor selecione o código de linguagem apropriado para os metadados.

(Adicionar a linguagem do standard de metadados de acordo com o ISO 639-3.)

Tópicos do PGD da FCT

- Informação sobre os Dados
- Documentação e Metadados
- Armazenamento e Segurança
 - Dados Pessoais e Direitos de Propriedade Intelectual
 - Partilha e Preservação a Longo Prazo
 - Responsabilidades e Recursos

- Dar preferência à utilização de um armazenamento robusto e gerido com apoio de **backup automático.**
- Não se recomenda o armazenamento de dados em computadores portáteis, discos rígidos autónomos, ou dispositivos de armazenamento externo, tais como dispositivos USB.
- Algumas instituições de investigação possuem drives de investigação em rede, que oferecem um espaço amplo de armazenamento e segurança de dados.



- Considerar a proteção de dados, especialmente se os dados forem sensíveis, por exemplo, se incluírem dados pessoais, informações politicamente sensíveis ou informações relacionadas com religião e saúde, segredos comerciais ou informações de segurança nacional;
- Explicitar o tipo de acesso aos dados durante o projeto, como é feito o controlo dos acessos e que política(s) de proteção de dados tem em vigor a instituição e/ou o arquivo/repositório.



Descrever onde é que os dados e os metadados serão armazenados durante o projeto.

Resposta

- ☐ Armazenamento dos dados na rede da instituição
- ☐ Outro (por favor especificar)

Projeto Quase

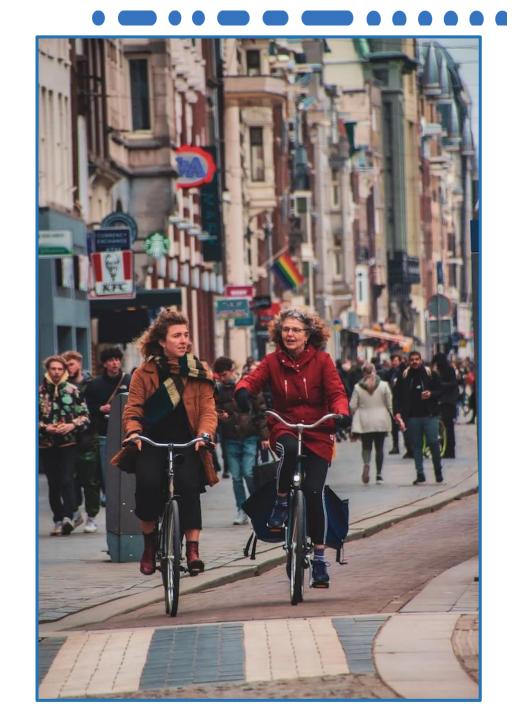
- ✓ Armazenamento dos dados na rede da instituição
- ☐ Outro (por favor especificar)

A recolha de dados terá início no mês de Janeiro de **2023** e final no mês de **Janeiro de 2025**.

O dispositivo tem capacidade estimada para armazenar os dados referentes a 5 eventos, sendo feita uma sincronização automática com o servidor de dados da instituição durante o carregamento do mesmo.

Ao serem depositados no servidor de dados da instituição, todos os dados são acompanhados de um manual de operações que define as condições de recolha dos dados.

O servidor de dados da instituição opera com uma infraestrutura redundante, sendo que os dados são replicados em duas instâncias do servidor de dados da instituição.



Existe uma política de salvaguarda de dados durante o projeto?

Resposta

- ☐ Sim
- □ Não
- ☐ Não sabe

Projeto Quase

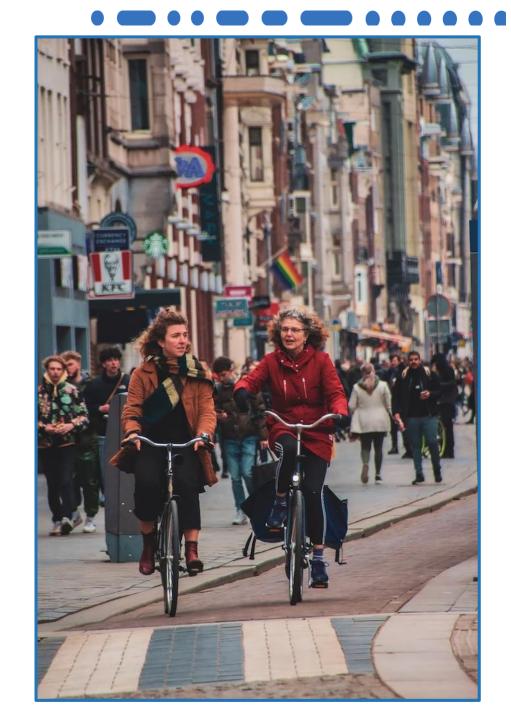
- ✓ Sim
 - Por favor descreva o tipo de backup
 - Por favor descreva a frequência do backup
- □ Não
- ☐ Não sabe

A recolha de dados terá início no mês de Janeiro de **2023** e final no mês de **Janeiro de 2025**.

O dispositivo tem capacidade estimada para armazenar os dados referentes a 5 eventos, sendo feita uma sincronização automática com o servidor de dados da instituição durante o carregamento do mesmo.

Ao serem depositados no servidor de dados da instituição, todos os dados são acompanhados de um manual de operações que define as condições de recolha dos dados.

O servidor de dados da instituição opera com uma infraestrutura redundante, sendo que os dados são replicados em duas instâncias do servidor de dados da instituição.



Proteção de Dados Sensíveis:

- Considere a proteção de dados, particularmente se os seus dados forem sensíveis.
- **Descreva os principais riscos** e a forma como estes serão **geridos**.
- Explique como os dados serão recuperados no caso de um incidente.
- Explique quem terá acesso aos dados durante a investigação e como é que será controlado esse acesso.
- Explique quais são as políticas institucionais de proteção de dados em vigor.



Como pretende assegurar a proteção e segurança dos dados sensíveis durante a investigação?

Resposta

- ☐ Não aplicável (não existem dados sensíveis)
- ☐ Serão aplicadas as medidas de segurança que estejam implementadas por defeito na rede da instituição
- ☐ Serão aplicadas medidas de segurança adicionais (por favor especificar)

Projeto Quase

- ☐ Não aplicável (não existem dados sensíveis)
- ✓ Serão aplicadas as medidas de segurança que estejam implementadas por defeito na rede da instituição
- ☐ Serão aplicadas medidas de segurança adicionais (por favor especificar)

Serão considerados os seguintes conjuntos de dados:

Identificação de Ciclistas

Tabela de identificação dos ciclistas

Formatos:

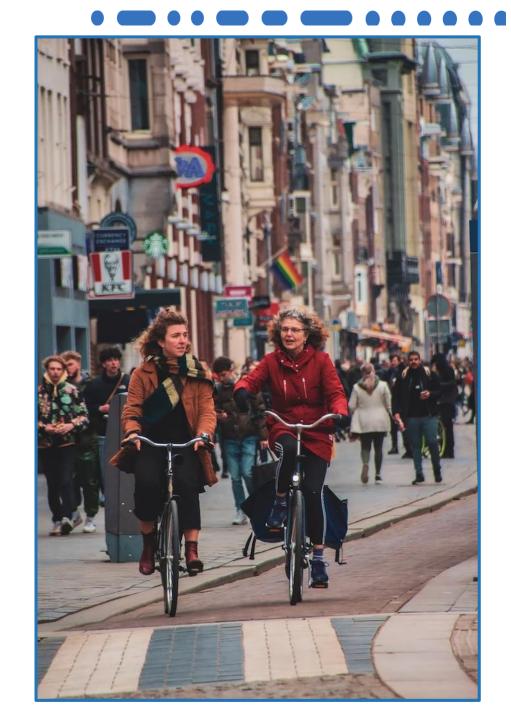
.CSV

Volume:

• < 1GB

Metadados:

- Nome do voluntário (String) Idade do voluntário
- (String) Número de cartão
- de cidadão (String) Identificador (ID atribuído ao voluntário)



Tópicos do PGD da FCT

- Informação sobre os Dados
- Documentação e Metadados
 - Armazenamento e Segurança
 - Dados Pessoais e Direitos de Propriedade Intelectual
 - Partilha e Preservação a Longo Prazo
 - Responsabilidades e Recursos

Dados pessoais

- **Informação pessoal** (nome, morada, cartão do cidadão, carta de condução, etc.).
- Informação sensível (dados pessoais que revelem a origem racial ou étnica; opiniões políticas e convicções religiosas ou filosóficas; filiação sindical; dados genéticos; dados biométricos; dados relacionados com a saúde; dados relativos à vida sexual ou orientação sexual da pessoa).
- Informação confidencial (acordada entre duas partes, como por exemplo, informação sobre o salário, ou outra informação que não esteja em domínio público).



https://amagno.co.uk/gdpr-legal-issues-protection-of-personal-data/22099/

PGD na prática

Vai processar e/ou armazenar dados pessoais durante o seu projeto?

Resposta

- ☐ Sim
- □ Não

Projeto Quase

- ✓ Sim
 - Como será assegurado o cumprimento da legislação e da regulamentação (institucional) sobre os dados pessoais?
- □ Não

Processamento e/ou armazenamento de dados pessoais

- Procure aconselhamento junto do gabinete de apoio especializado na sua Instituição
- Assegure o cumprimento com a lei de proteção de dados (RGPD) e que são cumpridos os seguintes requisitos:
 - Obter o consentimento informado para a preservação e/ou partilha dos dados pessoais
 - Considerar a anonimização ou pseudoanonimização de dados pessoais
 - Ponderar a encriptação que será considerada caso opte pela pseudo-anonimização
 - Estabelecer um procedimento de acesso aos dados pessoais para utilizadores autorizados



https://www.privacy365.eu/en/by-the-french-data-protection-authority-cnil-personal-data-anonymization/

Direitos de propriedade intelectual e propriedade dos dados

- **Direitos legais** associados aos dados:
 - Trade secret (informação proprietária ou confidencial)
 - Direitos morais e patrimoniais (Copyright)
 - Proteção jurídica de bases de dados
 - Patente



https://excelonip.com/an-overview-on-intellectual-property-rights-and-its-international-position/

PGD na prática

Como será gerida a propriedade dos dados e dos direitos de propriedade intelectual sobre os dados?

Projeto Quase

O Projeto Quase é um trabalho de investigação liderado pela **Universidade de Évora**, sendo **financiado** pela FCT através do grant FCT/2023/12345, com duração prevista de 2 anos.

Os dados gerados no âmbito do Projeto Quase são propriedade da Universidade de Évora. Sendo que estarão disponíveis para consulta mediante acesso controlado.

Dados Pessoais e Direitos de Propriedade Intelectual

Direitos de propriedade intelectual e propriedade dos dados

- Explicar quem será o proprietário dos dados, ou seja, quem terá os direitos de controlar o acesso aos mesmos.
- Deverá certificar-se que estão asseguradas todas as questões relativas ao direito de controlo do acesso aos dados, nomeadamente em projetos com vários parceiros e múltiplos proprietários de dados.
- Indicar se os direitos de propriedade intelectual (por exemplo, base de dados e/ou direitos) são afetados. Em caso afirmativo, explicar quais e como serão tratados.
- Indicar se existe alguma restrição à reutilização de dados de terceiros.

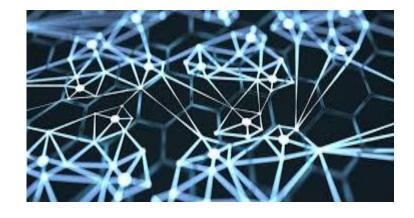


https://www.freepik.com/free-vector/abstract-illustration-intellectual-property-concept_10359462.htm#query=intellectual%20property&position=15&from_view=keyword

Tópicos do PGD da FCT

- Informação sobre os Dados
- Documentação e Metadados
- Armazenamento e Segurança
 - Dados Pessoais e Direitos de Propriedade Intelectual
 - Partilha e Preservação a Longo Prazo
 - Responsabilidades e Recursos

- Manter e preservar de forma fidedigna e confiável os dados de investigação;
- O hardware, software e até mesmo formatos de ficheiros tornar-se-ão desatualizados, devendo ser assegurada a preservação dos dados;
- Exige uma monitorização ativa da integridade dos dados e a antecipação de vários riscos;
- **Elementos chave:** metadados, proveniência, ferramentas ou software para reutilizar os dados.



Decisões importantes a tomar:

- Que dados serão preservados e por quanto tempo?
- Que dados estarão disponíveis para reutilização?
- Quando é que os dados estarão disponíveis para reutilização?
- Por quanto tempo estarão os dados disponíveis?



Como vai selecionar os dados para preservação de longo prazo?

Resposta

- ☐ Todos os dados resultantes do projeto serão preservados pelo menos durante 10 anos
- ☐ Outro (por favor especificar)

- ✓ Todos os dados resultantes do projeto serão preservados pelo menos durante 10 anos
- ☐ Outro (por favor especificar)

Que dados serão disponibilizados para reutilização?

Resposta

- ☐ Todos os dados resultantes do projeto serão disponibilizados
- ☐ Outro (por favor especificar)

- ✓ Todos os dados resultantes do projeto serão disponibilizados
- ☐ Outro (por favor especificar)

Quando estarão os dados disponíveis para reutilização e por quanto tempo os dados estarão disponíveis?

Resposta

- ☐ Os dados vão ficar disponíveis assim que forem arquivados no repositório
- ☐ Os dados vão ficar disponíveis assim que o artigo for publicado
- ☐ Os dados vão ficar disponíveis após a conclusão do projeto
- ☐ Os dados vão ficar disponíveis após a conclusão do projeto (com embargo)

- ☐ Os dados vão ficar disponíveis assim que forem arquivados no repositório
- ☐ Os dados vão ficar disponíveis assim que o artigo for publicado
- ✓ Os dados vão ficar disponíveis após a conclusão do projeto
- ☐ Os dados vão ficar disponíveis após a conclusão do projeto (com embargo)

Repositórios e licenças:

- Descrever em que repositório serão depositados os dados para preservação e reutilização;
- Definir a licença sob a qual esse depósito é efetuado;
- Descrever como irá ser disponibilizado o software necessário para a análise dos dados gerados no âmbito do projeto.





Repositórios confiáveis

Aconselha-se o cumprimento dos seguintes critérios mínimos de seleção:

- Fornecimento de identificadores únicos e persistentes;
- Utilização de normas de metadados que sejam amplamente aceites pela comunidade;
- Fornecimento de informação que esteja publicamente disponível;
- Permitir o acesso aos dados em condições específicas e seguindo protocolos de acesso abertos e normalizados;
- Disponibilização de informação sobre licenças e permissões;
- Garantia da preservação de dados e metadados.





ISO 16363 CERTIFICATION

Licenças

Disponibilize os seus dados de investigação sob uma licença apropriada, que defina os direitos de utilização dos mesmos.



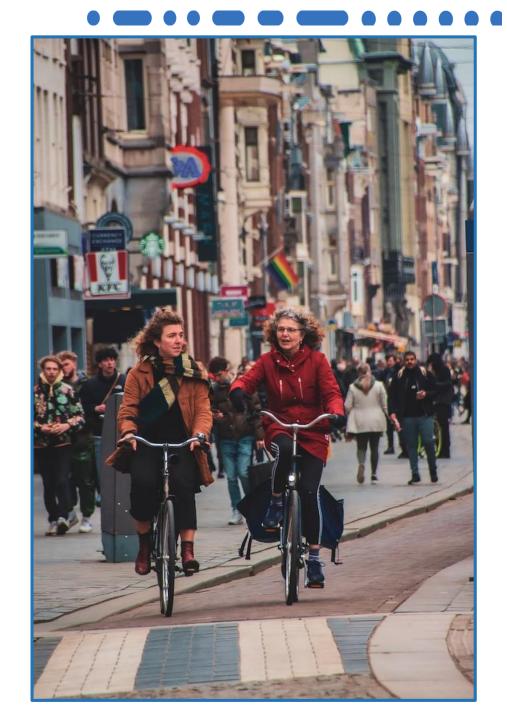
Escolha uma licença que:

- Assegure que a autoria e proveniência dos dados sejam corretamente atribuídas;
- Torne os termos de reutilização dos seus dados explícitos para o utilizador.

Em que repositório serão os dados arquivados e disponibilizados para reutilização?

- Serviço de Repositório de Dados de Investigação
- Baseado no software Dataverse
- Esforço conjunto da FCT | FCCN e INCD
- Sistema de PID (DOI) suportado pelo DataCite
- Licenças por defeito Creative Commons 4.0





Em que repositório serão os dados arquivados e disponibilizados para reutilização?

Por favor confirme

Resposta

- ☐ O repositório encontra-se registado no diretório de dados r3data
- ☐ O repositório encontra-se agregado no OpenAIRE
- ☐ O repositório é certificado
- ☐ Nenhuma das opções anteriores

- ✓ O repositório encontra-se registado no diretório de dados r3data
- ✓ O repositório encontra-se agregado no OpenAIRE
- ☐ O repositório é certificado
- ☐ Nenhuma das opções anteriores

O repositório suporta um sistema de identificadores persistentes?

Resposta

- ☐ Sim
- □ Não
- ☐ Desconhecido

- ✓ Sim
- □ Não
- □ Desconhecido

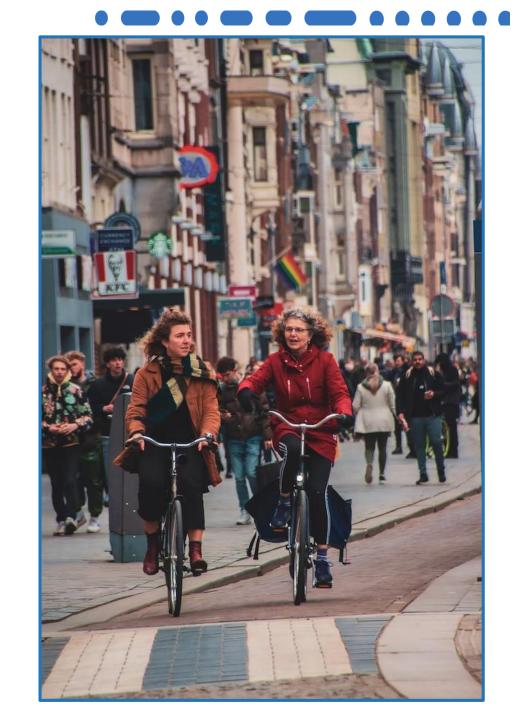
O repositório suporta um sistema de identificadores persistentes?

Sim

Por favor indique o sistema de PID

- Serviço de Repositório de Dados de Investigação
- Baseado no software Dataverse
- Esforço conjunto da FCT | FCCN e INCD
- Sistema de PID (DOI) suportado pelo DataCite
- Licenças por defeito Creative Commons
 4.0

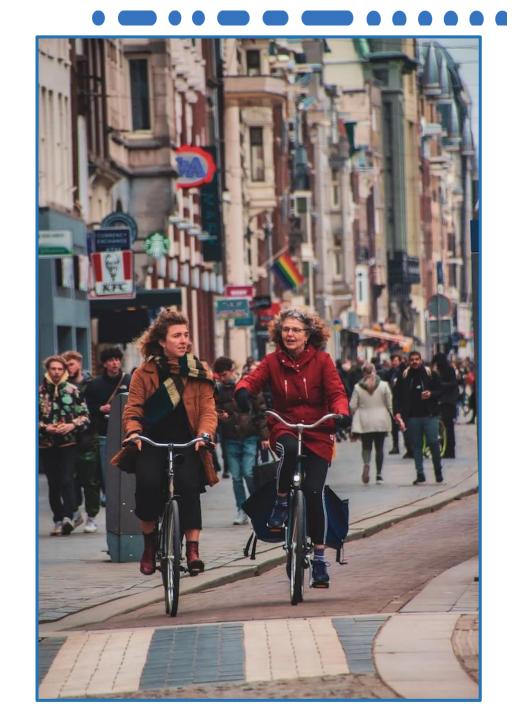




Sob que licença serão os dados arquivados e disponibilizados para reutilização?

- Serviço de Repositório de Dados de Investigação
- Baseado no software Dataverse
- Esforço conjunto da FCT | FCCN e INCD
- Sistema de PID (DOI) suportado pelo DataCite
- Licenças por defeito Creative Commons 4.0





Descreva a sua estratégia para disponibilizar as ferramentas e/ou software que sejam necessários para:

- o acesso
- **USO**
- reutilização dos dados.

A recolha de dados terá início no mês de **Junho de 2022** e final no mês de **Julho de 2022**.

O dispositivo tem capacidade estimada para armazenar os dados referentes a 5 eventos, sendo feita uma sincronização automática com o servidor de dados da instituição durante o carregamento do mesmo

Ao serem depositados no servidor de dados da instituição, todos os dados são acompanhados de um manual de operações que define as condições de



Tópicos do PGD da FCT

- Informação sobre os Dados
- Documentação e Metadados
 - Armazenamento e Segurança
 - Dados Pessoais e Direitos de Propriedade Intelectual
 - Partilha e Preservação a Longo Prazo
 - Responsabilidades e Recursos

O PGD engloba todas as fases do ciclo de vida dos dados de investigação e todos os conjuntos de dados resultantes da investigação. A responsabilidade e recursos são aspectos a ter em conta na criação de um PGD.

Apesar de não haver uma norma explícita quanto à responsabilidade e recursos de um PGD deve se ter presente as boas práticas.

A **responsabilidade** pela gestão dos dados deve ser partilhada por mais de um elemento da equipa de investigação.

Deve descrever as **funções** e **responsabilidades** das atividades relativas à gestão dos dados, tais como:

- recolha dos dados;
- produção de metadados;
- qualidade dos dados;
- armazenamento e cópia de segurança;
- arquivo dos dados;
- partilha dos dados.

Indicar os **recursos** necessários num PGD:

- Financeiros
- Partilha e Preservação-Repositórios confiáveis de dados de Investigação
- Eventuais recursos adicionais que podem advir da preparação dos dados para depósito ou para financiar encargos associados a repositórios.

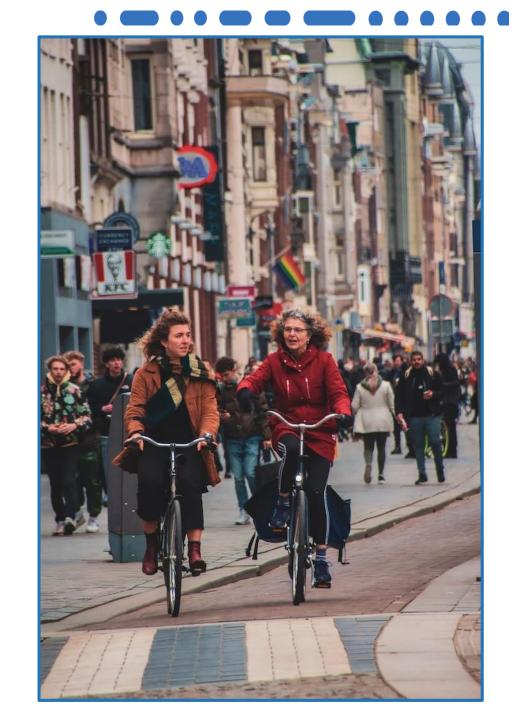
Quem (p.ex., função cargo e/ou instituição) será responsável pela gestão dos dados?

Quem (p.ex., função cargo e/ou instituição) será responsável pela gestão dos dados?

Por favor especifique

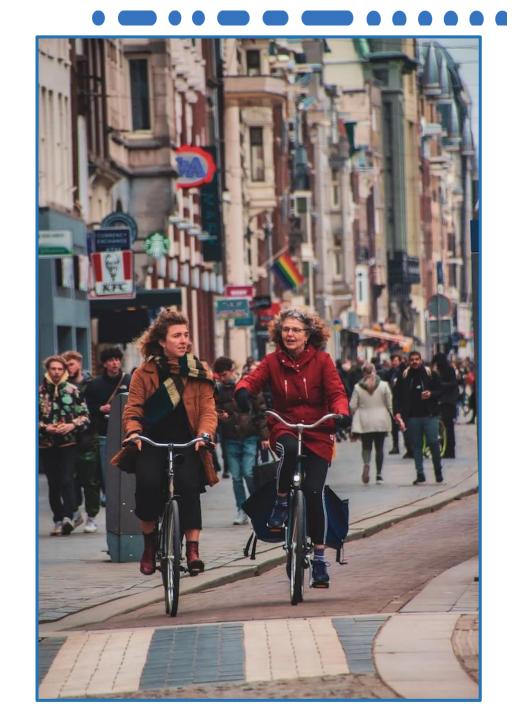
O **investigador principal** é a investigadora convidada Filomena Borba. Também na equipa estão presentes:

- André Vieira, João Cardoso e Filipa Pereira com tarefas de gestão de dados;
- Patrícia Miranda responsável pelo armazenamento durante a execução do projeto;
- Salima Rehemtula com tarefas de Data Protection Officer;
- Patrícia Dias de Melo responsável pela administração técnica do projeto.



Que recursos serão dedicados à gestão dos dados e à garantia de que os dados serão FAIR?





Perguntas e Respostas

OBRIGADO!

Oradores:

- João Cardoso
- Filomena Borba
- Filipa Pereira
- André Vieira
- Patrícia Miranda
- Salima Rehemtula
- Patrícia Dias de Melo