

Perguntas frequentes

Centro de Gestão de Dados de Investigação Interdisciplinar da Universidade de Lisboa
(iRe:Search)

Janeiro 2026
V1.1

Índice

Perguntas e Respostas (FAQ).....	3
Ciência Aberta e Dados de Investigação Abertos	3
Gestão e Preservação de Dados Abertos, Condicionais Legais e Regulamentares	9
Direitos de Autor e Direitos Conexos	12
Regulamento Geral de Proteção de Dados (RGPD)	16
Normas Internacionais e Boas Práticas	20
Ferramentas e Recursos Disponíveis.....	23
Ética e Transparência	26
Reprodutibilidade.....	30
Princípios de Gestão de Dados de Investigação e DMPs	33
Utilização de Inteligência Artificial (IA) em Dados de Investigação	36
Links e Recursos Essenciais para Ciência Aberta e GDI.....	39

Perguntas e Respostas (FAQ)

Ciência Aberta e Dados de Investigação Abertos

- O que é Ciência Aberta?
 - Movimento que visa tornar o processo de investigação (métodos, dados, publicações, software) acessível a todos.
- O que preconiza a Declaração de Barcelona sobre Informação de Investigação Aberta?
 - A Declaração de Barcelona exige que a informação sobre a investigação (metadados bibliográficos, referências, financiamento) seja aberta por omissão. O objetivo é reduzir a dependência de infraestruturas proprietárias fechadas para tomar decisões e realizar avaliações científicas, e garantir que os dados usados para avaliar investigadores sejam transparentes e auditáveis.
- O que muda com a estratégia "Towards Responsible Publishing" do cOAlition S em 2026?
 - A partir de 2025/2026, o cOAlition S (Plano S) reduziu o apoio financeiro aos acordos transformativos (pagamento de taxas em revistas híbridas). A prioridade passa a ser o apoio a modelos liderados pela comunidade (p.ex. Diamond Open Access) e a valorização de "preprints" e revisão por pares aberta, dissociando a validação científica do prestígio da revista.
- O que significa o princípio "tão aberto quanto possível, tão fechado quanto necessário"?
 - Resposta: Este é o princípio orientador da gestão de dados na União Europeia. Significa que os resultados da investigação devem ser partilhados publicamente por defeito, mas o acesso pode (e deve) ser restrito em casos de proteção de dados pessoais (RGPD), segurança nacional, direitos de propriedade intelectual (patentes em curso) ou interesses comerciais legítimos.

Fonte: [Horizon Europe Programme Guide \(p. 38\)](#)

- A Ciência Aberta é obrigatória em todos os projetos de investigação?
 - Não. A adoção de práticas de Ciência Aberta depende do enquadramento do projeto, do financiador e do tipo de resultados produzidos. No entanto, muitos financiadores nacionais e europeus exigem atualmente o acesso

aberto a publicações e a gestão adequada dos dados de investigação, tornando a Ciência Aberta um elemento cada vez mais central na prática científica. No Horizonte Europa, a integração de práticas como Open Peer Review ou Early Sharing (pré-registos) aumenta a pontuação na avaliação da metodologia.

- Qual o benefício de integrar a "Co-criação" e a "Ciência Cidadã" nas propostas do Horizonte Europa?
 - Além de aumentar o impacto social, projetos que envolvem ativamente stakeholders (indústria, cidadãos) desde o design experimental obtêm pontuações superiores no critério de "Metodologia de Ciência Aberta". Isso demonstra que os resultados não são apenas publicados, mas têm utilidade direta e imediata para a sociedade.

Fonte: [Horizon Europe Programme Guide](#)

- Qual a diferença entre Ciência Aberta, Open Innovation e Ciência Cidadã?
 - A Ciência Aberta centra-se na abertura e transparência do processo científico. A Open Innovation foca-se na colaboração entre academia, indústria e sociedade para inovação. A Citizen Science envolve cidadãos não académicos na recolha, análise ou validação de dados científicos. Estes conceitos são complementares e podem coexistir num mesmo projeto.
- A Ciência Aberta aplica-se também às Ciências Sociais e Humanidades?
 - Sim. Embora os tipos de dados e metodologias sejam diferentes, os princípios da Ciência Aberta — transparência, documentação adequada e partilha responsável — são aplicáveis a todas as áreas científicas, respeitando sempre os constrangimentos éticos, legais e culturais específicos.
- Como se monitoriza a Ciência Aberta?
 - Através de indicadores técnicos em dashboards (p.ex. EOSC Observatory) baseados nos princípios OSMI, que exigem infraestruturas abertas e dados de proveniência rastreáveis.
- O que são dados de investigação (Research Data)?
 - São registos factuais recolhidos, observados ou gerados durante o processo de investigação, que servem de base para a análise e validação de resultados científicos.

- É obrigatório publicar todos os dados recolhidos durante a investigação?
 - Não. Dados sem qualidade (p.ex. medições laboratoriais falhadas) ou que não sejam úteis para a própria investigação não têm obrigatoriedade de publicação. No entanto, se os dados forem válidos, mesmo que não sejam utilizados no estudo atual para a própria investigação, devem ser publicados se puderem ser úteis para terceiros no futuro.

- O que são conjuntos de dados (Dataset)?
 - Coleção estruturada de dados relacionados, geralmente organizada em formato tabular, numérico, textual, visual ou multimédia, e acompanhada dos respetivos metadados.

- Como citar corretamente um dataset?
 - Deve incluir: Autor(es), Ano, Título, Versão, Repositório e o Identificador Persistente (DOI).

- O que são metadados (Metadata)?
 - São o "Rótulo" dos dados: Informação descritiva sobre os dados que permite compreender, localizar e reutilizar um conjunto de dados.

- O que é um repositório de dados (Data Repository)?
 - É uma plataforma digital segura onde os conjuntos de dados e metadados são depositados, preservados e tornados acessíveis ao público ou a grupos autorizados. Deve garantir a preservação a longo prazo e a interoperabilidade.

- Quem é o proprietário (Owner) legal dos dados de investigação?
 - Geralmente, a instituição de acolhimento detém a titularidade dos dados produzidos no âmbito de contratos de trabalho ou financiamento, embora o investigador detenha os direitos morais e de autoria científica.

- Qual a diferença entre Data Owner, Data Controller e Data Processor?
 - Data Owner: entidade com responsabilidade institucional.
 - Data Controller: define finalidades e meios do tratamento (RGPD).
 - Data Processor: trata dados em nome do controller.

- O que faz um Data Steward?
 - É o profissional (ou ponto de contacto) que atua como ponte entre a TI, o jurídico e os investigadores, garantindo a qualidade dos metadados e a implementação prática dos princípios FAIR no dia a dia do projeto.

- O que é governação de dados em investigação?
 - Conjunto de políticas, papéis e responsabilidades que asseguram uma gestão legal, ética e sustentável dos dados.
- Qual o papel dos serviços institucionais?
 - Bibliotecas, IT, jurídico e data stewards apoiam curadoria, licenciamento, conformidade e preservação.
- Como posso citar corretamente um conjunto de dados (dataset) de outro investigador?
 - A citação de dados deve ser tratada com o mesmo rigor que a citação de um artigo. Deve incluir: Autor(es), Ano de Publicação, Título do Dataset, Versão, Repositório e o Identificador Persistente (geralmente um DOI).

Exemplo ilustrativo (o DOI não é real): Silva, J. (2024). Dados de temperatura oceânica (V1.2) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.12345>

- Como é feita a monitorização da Ciência Aberta a nível institucional e europeu?
 - A monitorização deixou de ser apenas qualitativa para se basear em indicadores técnicos via dashboards (p.ex. EOSC Observatory, PathOS). Estes sistemas seguem os "Princípios de Monitorização da Ciência Aberta" (OSMI), que exigem o uso de infraestruturas abertas e dados de proveniência rastreáveis para evitar métricas proprietárias fechadas.

(Fonte:<https://open-science-monitoring.org/>)

- Como a Ciência Aberta influencia a avaliação de investigadores e instituições?
 - Cada vez mais, a avaliação científica na UE privilegia a diversidade de outputs (dados, software, pré-registos, revisão aberta) e o impacto social, em linha com DORA e a Reforma da Avaliação da Investigação (CoARA), o que reduz a dependência de métricas baseadas em revistas.
- O que distingue FAIR de Acesso Aberto (Open Access)?
 - O FAIR vai além do Acesso Aberto, concentrando-se na qualidade técnica dos dados para máquinas, enquanto o Acesso Aberto garante a leitura gratuita da publicação científica.
- Qual é o foco do FAIR?
 - O foco principal é a qualidade técnica da informação e visa garantir que os dados de investigação possam ser descobertos e reutilizados por humanos e máquinas.

- Dados FAIR significam dados Abertos (Open Data)?
 - Não. Dados FAIR podem ser fechados ou restritos (p.ex. dados médicos sensíveis). O princípio é "tão aberto quanto possível, tão fechado quanto necessário", desde que os metadados (a descrição dos dados) sejam públicos e acessíveis. O FAIR exige boa gestão, não necessariamente abertura total.
- Como avaliar se os dados são FAIR?
 - Utilizando ferramentas reconhecidas como FAIR-Aware e F-UJI.
- Os dados têm de ser 100% FAIR?
 - Não. FAIR é um contínuo e deve ser aplicado de forma proporcional.
- Como posso saber se um repositório é confiável para preservação a longo prazo (Princípios TRUST)?
 - Enquanto os princípios FAIR se aplicam aos dados, os princípios **TRUST** aplicam-se aos repositórios: **T**ransparência, **R**esponsabilidade, **U**suário (foco no), **S**ustentabilidade e **T**ecnologia. Recomenda-se a escolha de repositórios certificados (p.ex. CoreTrustSeal) que cumpram estes requisitos para garantir que os dados não desaparecem ou se corrompem.
- O que são Vocabulários Controlados e por que importam?
 - São listas padronizadas de termos (p.ex. MeSH, Getty Thesaurus) que garantem que diferentes computadores "falem a mesma língua", essenciais para cumprir o "I" (Interoperabilidade) do FAIR.
- Quando deve ser feita a avaliação FAIR?
 - Idealmente durante o projeto (FAIR-by-design) e novamente antes do depósito final dos dados no repositório.
- A avaliação FAIR substitui a revisão científica?
 - Não. A avaliação FAIR verifica a qualidade técnica e a reutilização dos dados, não a validade científica dos resultados.
- A avaliação FAIR é usada em projetos financiados?
 - Sim. Em projetos Horizonte Europa e EOSC, a avaliação FAIR é utilizada como instrumento de monitorização técnica da qualidade dos dados e da conformidade dos PGDs, complementando a avaliação científica tradicional.

- Os princípios FAIR significam que os dados têm de estar em acesso aberto total?
 - Não necessariamente. O princípio é "tão aberto quanto possível, tão fechado quanto necessário". É possível cumprir os princípios FAIR mantendo os dados restritos (p.ex. dados sensíveis), desde que os metadados sejam públicos e indiquem claramente as condições e o protocolo de acesso controlado.

- Quais são Princípios CARE?
 - Os quatro componentes dos Princípios CARE são: Collective Benefit (Benefício Coletivo; Authority to Control (Autoridade para Controlar; Responsibility (Responsabilidade) e Ethics (Ética).

- Qual é a principal responsabilidade ética (R) [dos Princípios CARE]?
 - A Responsabilidade (Responsibility) implica transparência no uso dos dados e perante quem os forneceu (por vezes designados como donos dos dados).

- O que é o EOSC (European Open Science Cloud)?
 - É uma plataforma federada Europeia que fornece aos investigadores e cientistas cidadãos acesso a dados, publicações, software e serviços interoperáveis e reutilizáveis. Oferece ainda serviços chave como File Sync & Share e Interactive Notebooks.

- A Ciência Aberta implica sempre custos adicionais para o projeto?
 - Não necessariamente. Muitas práticas de Ciência Aberta têm custos marginais reduzidos e os custos elegíveis podem ser incluídos em financiamentos como Horizonte Europa e FCT.

Gestão e Preservação de Dados Abertos, Condicionalismos Legais e Regulamentares

- Qual é o requisito mínimo de retenção de dados?
 - 10 anos, ou justificar a aplicação de um período diferente.
- O que acontece aos dados após o fim do projeto?
 - Devem permanecer preservados no repositório, de acordo com o período de retenção definido.
- Quem suporta os custos após o projeto?
 - Depende do modelo institucional e do repositório, frequentemente absorvidos por infraestruturas públicas.
- Porque é que a preservação de dados é uma responsabilidade institucional e não apenas do projeto?
 - Porque a preservação a longo prazo exige infraestruturas, políticas e financiamento sustentáveis, que ultrapassam a duração de projetos individuais. Por isso, os financiadores recomendam repositórios institucionais, nacionais ou certificados.
- Quais são as justificações válidas para restringir o acesso aos dados (mantê-los fechados)?
 - As exceções devem ser justificadas de forma clara e incluem: valor comercial (segredos comerciais, proteção de PI), proteção de dados e privacidade (garantir a conformidade com o RGPD), preocupações de segurança (projetos de defesa), e interesses estratégicos da UE.
- O que é um Repositório Confiável (Trusted Repository)?
 - Um Repositório Confiável (*Trusted Repository*) é uma plataforma digital que cumpre padrões internacionais rigorosos para garantir a integridade, a autenticidade e a preservação a longo prazo dos conteúdos que armazena, incluindo dados, publicações e *software*.
- O que é a regra 3-2-1 para cópias de segurança?
 - É uma estratégia de preservação que recomenda ter 3 cópias dos dados, em 2 suportes diferentes (p.ex. disco externo e nuvem), com 1 das cópias num local físico distinto (fora da instituição/casa).

- Qual o padrão de backup recomendado?
 - A Regra 3-2-1 (3 cópias, 2 tipos de media diferentes, 1 cópia off-site). Deve-se preferir armazenamento em rede em vez de discos locais
- Onde armazenar os dados durante um projeto de investigação?
 - Em armazenamento seguro e em rede (servidores institucionais ou nuvem aprovada), evitando discos locais ou pen drives.
- Como escolher entre um repositório institucional, disciplinar ou genérico?
 - Sempre que exista um repositório disciplinar reconhecido, essa opção deve ser privilegiada. Na ausência deste, recomenda-se o uso de repositórios institucionais ou, em alternativa, repositórios genéricos certificados. A escolha deve considerar requisitos do financiador, interoperabilidade, preservação e visibilidade.
- O que acontece se o repositório deixar de existir?
 - Repositórios certificados e integrados em infraestruturas de preservação implementam estratégias para garantir acesso a longo prazo. A utilização de repositórios reconhecidos e certificados reduz significativamente o risco de perda de dados.
- Posso depositar dados em mais do que um repositório?
 - Sim, desde que exista coerência entre as versões depositadas e que sejam claramente indicadas as versões oficiais e os identificadores persistentes associados.
- Como gerir dados de sensores que são enviados automaticamente para uma cloud proprietária?
 - Este fluxo deve ser descrito no Plano de Gestão de Dados (PGD). O investigador deve descarregar os dados dessa cloud, processá-los, adicionar metadados e disponibilizá-los num formato aberto para garantir a preservação a longo prazo, funcionando a cloud apenas como uma etapa do sistema de recolha de dados.
- Os custos de preservação de dados após o final do projeto são elegíveis para financiamento?
 - Durante a vigência do projeto, custos de curadoria e armazenamento são elegíveis se previstos na proposta/candidatura. No entanto, faturas após o fim do projeto não são em regra aceites pelos financiadores; a preservação pelos 10 anos exigidos deve ser assegurada por

financiamento institucional, receitas de exploração ou projetos de continuidade.

- Qual a diferença entre "Armazenamento de Dados" e "Preservação de Dados"?
 - O armazenamento refere-se a guardar ficheiros para uso imediato durante o projeto (p.ex. Dropbox, disco rígido). A preservação é um processo ativo a longo prazo (10+ anos) que garante que os dados continuam legíveis apesar da obsolescência tecnológica, envolvendo a migração de formatos (p.ex. converter .xls para .csv) e a verificação regular de integridade.

Fonte: [Digital Curation Centre \(DCC\) - Curation Lifecycle](#)

- O que é o Identificador de Atividade de Investigação (RAiD) e como se distingue do DOI?
 - Enquanto o DOI identifica um objeto digital estático (como um *dataset* ou artigo), o RAiD (ISO 23527:2022) identifica o **projeto de investigação** como um todo. Ele funciona como um "envelope" que liga pessoas (ORCID), organizações (ROR) e *outputs* (DOI), permitindo rastrear a evolução do projeto e a sua proveniência ao longo do tempo.

(Fonte: <https://www.raid.org/>)

- Como garantir a integridade técnica durante a transferência de grandes volumes de dados para preservação?
 - Recomenda-se o uso da especificação **BagIt** (da Library of Congress). Um "bag" é uma estrutura de pastas que inclui manifestos de *checksum* (assinaturas digitais) para garantir que nenhum ficheiro foi corrompido ou alterado durante a migração entre sistemas ou para repositórios como o Archivematica.

(Fonte: <https://www.archivematica.org/en/docs/archivematica-1.8/user-manual/transfer/bags/>)

Direitos de Autor e Direitos Conexos

- O que são Licenças Abertas?
 - São instrumentos legais que definem como os dados e publicações podem ser usados, partilhados e adaptados.
- O que distingue CC BY de CC0?
 - CC BY exige atribuição (dar crédito); CC0 dedica o trabalho ao Domínio Público, o que maximiza a reutilização por máquinas, sem obrigação de atribuir a autor(es).
- Qual é a licença CC mais recomendada para dados abertos?
 - CC BY 4.0 ou CC0. O CC0 é obrigatório para metadados e recomendado para maximizar a reutilização por máquinas.
- O que significa a licença MIT?
 - É uma licença Open Source para software que permite usar, aceder, modificar e partilhar o código fonte.
- Os dados brutos são protegidos por Direitos de Autor?
 - Não, mas uma base de dados pode ser protegida por Direito Sui Generis.
- O que é a Estratégia de Retenção de Direitos (RRS)?
 - Uma cláusula que permite aos autores referem direitos para depositar o Manuscrito Aceite do Autor (AAM) imediatamente (via verde) sob licença CC BY 4.0.
- Posso mudar a licença dos dados depois de publicados?
 - Em regra, não. Uma vez atribuída uma licença aberta e reutilizada por terceiros, a sua alteração não invalida os direitos já concedidos. Por isso, a escolha da licença deve ser ponderada antes da publicação.
- Qual a diferença prática entre CC BY e CC BY-SA?
 - Ambas permitem reutilização com atribuição. A CC BY-SA exige que obras derivadas sejam distribuídas sob a mesma licença, enquanto a CC BY não impõe essa obrigação.

- O que devo fazer se a editora não aceitar a cláusula de retenção de direitos?
 - Deve procurar revistas compatíveis com os requisitos de acesso aberto ou recorrer a ferramentas como o Journal Checker Tool. Deve ainda contactar os serviços de apoio à investigação da instituição.

- Qual a licença mais recomendada para dados?
 - CC BY 4.0 ou CC0. A CC0 é obrigatória para metadados e maximiza a reutilização por máquinas.

- Posso publicar um artigo em Acesso Aberto se a revista for "Híbrida"?
 - Sim, mas deve ter atenção aos custos. Uma revista híbrida é uma revista sob subscrição que permite abrir artigos individuais mediante o pagamento de uma taxa (APC). Nota importante: O Plano S (financiadores europeus) geralmente não financia APCs em revistas híbridas, a menos que estas façam parte de um "Acordo Transformativo" com a sua instituição (ULisboa).

- O que é o Plano S?
 - O Plano S é uma iniciativa lançada pela cOAlition S (um consórcio de agências financiadoras de ciência, incluindo a Comissão Europeia e, em Portugal, frequentemente alinhada com as diretrizes da FCT). O objetivo central é que toda a investigação financiada por fundos públicos deve ser publicada em Acesso Aberto imediato.

Link: [cOAlition S - Transformative Agreements](#)

- Quais as regras do Plano S em 2026?
 - Imediatismo: Nada de "períodos de embargo" (esperar 6 ou 12 meses para liberar o artigo).
 - Licenciamento: Exige-se geralmente uma licença CC BY (que permite a partilha e reutilização máxima).
 - Direitos do Autor: O autor deve reter os seus direitos de propriedade intelectual, não os cedendo totalmente à editora.

- O que é APC (Article Processing Charge)?
 - O APC (Taxa de Processamento de Artigo) é o valor cobrado por uma editora para cobrir os custos de publicação e tornar o seu artigo disponível em Acesso Aberto (Open Access).

- Como funciona o APC?
 - No modelo tradicional, o leitor (ou a biblioteca da universidade) paga uma assinatura para ler. No modelo de Acesso Aberto com APC, o autor (ou quem financia a pesquisa) paga uma taxa única, e o artigo fica gratuito para qualquer pessoa no mundo ler e baixar.

- Quais os custos a considerar em 2026?
 - Em 2026, esses valores variam muito, podendo ir de 500€ a mais de 5.000€, dependendo do prestígio da revista.

- Como posso licenciar hardware (equipamento) desenvolvido no âmbito da investigação científica?
 - Para hardware, as licenças de software ou CC não são ideais. Deve-se usar a CERN Open Hardware License (CERN-OHL v2), que possui três variantes: S (fortemente recíproca/copyleft), W (fracamente recíproca) e P (permissiva). Esta licença protege especificamente os designs e a fabricação física.

(Fonte: <https://ohwr.org/licences/>)

- O que são os "Direitos de Publicação Secundária" e como protegem o investigador?
 - É um direito legal (já implementado em vários países da UE via Knowledge Rights 21) que permite ao investigador republicar o seu trabalho em acesso aberto, independentemente de contratos assinados com editoras, desde que a investigação tenha financiamento público. Serve como uma salvaguarda jurídica para a "Via Verde" do Acesso Aberto.

(Fonte: <https://www.knowledgerights21.org/>)

- Os mesmos direitos e licenças aplicam-se a dados, software e publicações?
 - Não. Dados, software e publicações estão sujeitos a regimes jurídicos distintos e devem ser licenciados separadamente (p.ex. CC para dados/publicações; licenças open source para software; CERN-OHL para hardware).

- Qual a diferença entre licenciar dados com CC BY e ODbL?
 - As licenças CC focam-se no conteúdo criativo; a ODbL é desenhada especificamente para bases de dados, cobrindo direitos sobre a compilação e estrutura que as CC por vezes não detalham exaustivamente.

- É permitido atrasar a publicação científica para proteger uma patente?
 - Sim. Embora os financiadores exijam acesso aberto, eles aceitam um atraso na publicação para permitir a formalização de direitos de propriedade intelectual. Esta possibilidade deve estar prevista no PGD e não deve ocorrer qualquer divulgação pública (conferências ou preprints) antes do depósito da patente, para não invalidar a "novidade absoluta".

- Como garantir a citação correta de dados quando estes são processados por algoritmos ou máquinas?
 - Recomenda-se a utilização de infraestruturas que permitam a rastreabilidade (p.ex. OBIS na biodiversidade), onde a atribuição de um DOI aos dados originais permite que a citação seja mapeada e mantida mesmo após a integração em grandes conjuntos de dados processados automaticamente.

Regulamento Geral de Proteção de Dados (RGPD)

- O RGPD é incompatível com Ciência Aberta?
 - Não. O RGPD não proíbe a partilha de dados, mas exige que esta seja feita de forma lícita, ética e proporcional. A Ciência Aberta promove precisamente uma partilha responsável (“as open as possible, as closed as necessary”).
- A anonimização é sempre necessária para partilhar dados pessoais?
 - Sim, é necessária se a partilha for aberta (dados anónimos não são abrangidos pela RGPD). Se a anonimização robusta não for possível, o acesso deve ser controlado (managed access) via Acordo de Partilha de Dados (DSA).
- Existe alguma ferramenta recomendada para anonimização de dados?
 - Sim, o Amnesia (do OpenAIRE) é a ferramenta de referência que permite remover identificadores de datasets, gerando uma versão estatisticamente útil, mas anónima e protegendo dados pessoais.
- Devo usar o AMNESIA para dados confidenciais?
 - Sim, o uso online é apenas para demonstração, sendo recomendada a utilização local (on-premise) para maior segurança dos dados sensíveis.
- O que é Data Protection by Design?
 - É a exigência de integrar a proteção de dados desde a conceção do projeto.
- Quando é obrigatória a DPIA?
 - A DPIA é exigida pelo Regulamento Geral de Proteção de Dados (RGPD) quando o tratamento/partilha de dados pessoais possa resultar num risco elevado para os direitos e liberdades dos titulares dos dados, como saúde, genéticos, etc., ou utilização de novas tecnologias (IA, profiling).
- Anonimização vs. Pseudonimização?
 - A Anonimização Robusta retira os dados do âmbito do RGPD (permitindo a partilha aberta), enquanto a Pseudonimização é reversível (mantendo os dados sob o RGPD).

- O que são dados sintéticos?
 - São dados gerados artificialmente para mimetizar as propriedades estatísticas de um dataset real sem conter informações de indivíduos reais, garantindo a Privacidade por Design.

- O que é Privacidade Diferencial?
 - É um padrão matemático que garante que a inclusão ou exclusão de um indivíduo num dataset não altera os resultados de uma consulta, sendo ideal para Big Data e IA.

- Quais são os direitos dos titulares dos dados?
 - Sumariamente: direito de acesso, retificação, apagamento, limitação do tratamento, portabilidade, oposição a decisões automatizadas (ver RGPD para referência completa).

- Onde se regulam as Transferências Internacionais?
 - A transferência de dados pessoais para fora do EEE é regulada por legislação nacional e exige conformidade rigorosa.

- Como devo proceder para partilhar dados protegidos pelo RGPD?
 - Se a anonimização robusta não for possível, os dados não podem ter partilha aberta. Deve optar-se pelo acesso controlado (managed access) através de um Acordo de Partilha de Dados (DSA).

- Dados anonimizados podem voltar a ser considerados dados pessoais?
 - Sim, se existir risco de reidentificação devido à combinação com outras fontes de dados. A anonimização deve ser avaliada periodicamente, considerando o contexto tecnológico e científico.

- Quem é responsável pelo RGPD num consórcio internacional?
 - Cada parceiro é responsável pelo tratamento de dados sob o seu controlo, devendo existir acordos claros que definam responsabilidades, papéis e fluxos de dados.

- O consentimento é sempre a base legal mais adequada?
 - Não necessariamente. Em investigação científica, outras bases legais, como p.ex. o interesse público, podem ser mais adequadas. A escolha deve ser feita caso a caso.

- A recolha de dados anónimos requer o preenchimento de uma Avaliação de Impacto (AIPD)?
 - Se os dados forem verdadeiramente anónimos na origem (sem qualquer possibilidade de reidentificação), o RGPD não se aplica. No entanto, se os dados forem apenas "pseudonimizados" (existe uma chave que liga o código ao nome), o RGPD aplica-se e poderá ser necessária uma AIPD se o risco para os titulares for elevado.

Fonte: [Comissão Nacional de Proteção de Dados \(CNPD\)](#)

- O que são dados sintéticos e quando devem ser usados em vez de dados reais anonimizados?
 - Dados sintéticos são gerados artificialmente (via modelos como o Synthetic Data Vault - SDV) para mimetizar as propriedades estatísticas de um dataset original sem conter informações de indivíduos reais. Devem ser usados quando o risco de reidentificação é elevado ou quando é necessário partilhar utilidade analítica sem expor dados sensíveis, garantindo privacidade por design.

(Fonte: <https://docs.sdv.dev/sdv/>)

- O que é a "Privacidade Diferencial" em gestão de dados?
 - É um padrão matemático rigoroso que garante que a presença ou ausência de um único indivíduo num conjunto de dados não altera significativamente os resultados de uma consulta. É a técnica mais avançada para anonimização de Big Data e modelos de IA.

(Fonte: <https://docs.sdv.dev/sdv/explore/sdv-bundles/differential-privacy>)

- O que é o "Altruísmo de Dados" no contexto do Data Governance Act (DGA)?
 - O DGA estabelece um quadro legal para o "Altruísmo de Dados", permitindo que indivíduos e empresas disponibilizem voluntariamente dados (mesmo sensíveis, com consentimento) para fins de interesse geral, como a investigação científica. Estes dados são geridos por organizações de altruísmo de dados reconhecidas e registadas na UE, facilitando a recolha segura para investigadores.
- O RGPD aplica-se a dados recolhidos fora da União Europeia?
 - Dados recolhidos fora da UE estão sujeitos às leis locais do país de origem. Contudo, se esses dados forem transferidos para um servidor localizado na Europa, passam a estar sujeitos ao RGPD. A boa prática recomenda seguir o padrão mais restritivo para salvaguardar a instituição.

- Trabalhar com documentos públicos que identificam figuras públicas exige anonimização?
 - Pelo princípio da cautela, se o documento original não contiver uma declaração explícita de conformidade com o RGPD ou for anterior ao regulamento, é preferível anonimizar a informação na publicação da investigação (p.ex. "Participante X").

Normas Internacionais e Boas Práticas

- Porque são importantes as normas internacionais em projetos colaborativos?
 - Porque garantem interoperabilidade técnica, semântica e legal entre países, instituições e disciplinas, permitindo que dados e software sejam reutilizados globalmente.
- Quais PIDs são essenciais para o F de Findable?
 - O uso de DOI (para o dataset) e ORCID (para o autor) é fundamental.
- Por que razão o PGD exige PIDs?
 - Os PIDs garantem que os dados são Findable (F de FAIR) e citáveis, contribuindo para a integridade e rastreabilidade.
- O que é o ORCID?
 - É o Identificador Persistente para autores, essencial para ligar de forma inequívoca o investigador aos seus outputs.
- O que são os metadados e porque são vitais para os princípios FAIR?
 - Metadados são "dados sobre dados" (p.ex. data de recolha, autor, metodologia). Sem metadados padronizados, os dados são inúteis para outros investigadores porque não podem ser encontrados (Findable) ou interpretados (Reusable). Pense nos metadados como o rótulo de uma lata de comida; sem ele, não sabe o que está dentro nem se ainda é seguro consumir.

(Fonte: [GO FAIR - FAIR Principles Resources](#))

- Onde posso encontrar o standard de metadados para a minha área?
 - O Fairsharing.org ajuda a identificar padrões de metadados disciplinares (como ISO 19115 para dados geográficos).
- O que é Dublin Core?
 - É um conjunto de 15 elementos essenciais de metadados que todos os recursos digitais devem ter para serem localizáveis.
- Quais os 15 elementos [do Dublin Core]?
 - Concentram-se em três aspetos principais: Conteúdo (Título, Tópico), Propriedade Intelectual (Criador, Direitos, Editor) e Instanciação (Data, Formato, Identificador - DOI).

- Por que usar CRediT?
 - Porque garante a transparência e a responsabilidade na autoria, assegurando que todas as contribuições, incluindo as relacionadas com Dados e Software, são reconhecidas.
- Como o sistema CRediT ajuda no reconhecimento de contribuições técnicas além da autoria?
 - A taxonomia CRediT permite detalhar 14 funções específicas (p.ex. *Data Curation, Software, Validation*). Isto é vital para investigadores de carreira técnica ou *data stewards* cujas contribuições para a GDI são muitas vezes invisibilizadas em listas de autores tradicionais.

(Fonte:<https://credit.niso.org/>)

- O SWHID é mesmo necessário para o software, em vez de um DOI?
 - As soluções desenvolvidas para dados não são apropriadas para software devido ao seu ciclo de vida complexo e à sua natureza executável.
- Como posso garantir que o meu software é citável adequadamente?
 - Deve utilizar identificadores específicos como o SWHID para o código-fonte, uma vez que os DOIs tradicionais para dados podem não capturar a natureza executável e as versões complexas do software.
- É obrigatório usar Dublin Core se existir um standard disciplinar?
 - Não. Sempre que exista um standard disciplinar reconhecido, este deve ser preferido. O Dublin Core é frequentemente usado como camada mínima de interoperabilidade.
- Qual a diferença entre DOI e SWHID?
 - O DOI identifica objetos digitais publicados, enquanto o SWHID identifica de forma persistente código-fonte arquivado no Software Heritage.
- Os identificadores persistentes devem constar no texto do artigo?
 - Sim. Sempre que possível, devem ser incluídos tanto nos metadados como no texto do artigo para facilitar a citação e a rastreabilidade.
- O que é o padrão OBO identificado nas estatísticas de ontologias?
 - Refere-se à Open Biological and Biomedical Ontology, um padrão específico para as ciências biológicas e biomédicas que estrutura os termos e relações nestas áreas.

- Como descrever o fluxo de trabalho (workflow) em áreas não tecnológicas, como o Direito?
 - O workflow deve descrever p.ex. o passo a passo da investigação: o processo de recolha de jurisprudência em diferentes jurisdições, os critérios de seleção da informação relevante e as fases de análise. Isto permite que outros investigadores repliquem os métodos.

- Onde posso encontrar vocabulários controlados e padrões de metadados específicos para a minha área científica?
 - O recurso central é o FAIRsharing.org, que mapeia a interligação entre padrões de metadados, bases de dados e políticas por domínio científico (p.ex. Darwin Core para biologia, DDI para ciências sociais).

(Fonte: <https://fairsharing.org/>)

Ferramentas e Recursos Disponíveis

- Como escolher ferramentas de GDI (Gestão de Dados de Investigação) adequadas ao meu projeto?
 - A escolha deve basear-se nos requisitos do financiador, tipo e volume de dados, necessidades de interoperabilidade, certificação do repositório e apoio institucional disponível.
- O que é um maDMP?
 - Um machine-actionable DMP (PGD legível por máquina) é um ficheiro estruturado (JSON/XML) que permite a automatização de tarefas de gestão de dados e a pré-população de metadados em repositórios.
- O que faz o Binder?
 - Transforma repositórios de código (como Jupyter Notebooks) em ambientes computacionais executáveis no browser, garantindo a reprodutibilidade imediata.
- Qual é a capacidade máxima de upload no Zenodo?
 - Permite uploads até 50 GB por registo.
- Zenodo é adequado para depositar software?
 - Sim, possui funcionalidades para a preservação de software e integração com plataformas como o GitHub.
- Qual o papel do OpenAIRE?
 - Assegura a interoperabilidade da produção científica com repositórios nacionais, ORCID e a EOSC.
- O que é o RCAAP?
 - A Rede de Repositórios Científicos de Acesso Aberto de Portugal, onde todas as publicações científicas nacionais, p.ex. da ULisboa, devem ser depositadas.
- Qual o modelo PGD preferencial para a FCT?
 - O modelo FCT (Anexo II) encontra-se disponível na ferramenta ARGOS.

- O que é o Open Research Europe?
 - É uma plataforma de publicação open peer-reviewed para investigação financiada pela UE, que permite a publicação imediata do artigo como pre-print (CC-BY), seguida de open peer review.

- Qual é o benefício de usar ferramentas que suportam APIs?
 - Ferramentas com acesso via API permitem a automação de fluxos de trabalho, como o upload automatizado de grandes volumes de dados de sensores diretamente para um repositório (p.ex. Zenodo ou OSF), reduzindo erros manuais e promovendo um ecossistema de investigação conectado.

- Qual a diferença entre Zenodo, Figshare e repositórios institucionais?
 - Zenodo e Figshare são repositórios genéricos, enquanto os repositórios institucionais estão integrados nas políticas e serviços da instituição. A escolha depende do contexto e dos requisitos aplicáveis.

- Quando devo usar um caderno de laboratório eletrónico (ELN)?
 - Sempre que seja necessária rastreabilidade, colaboração e controlo de versões, especialmente em ambientes laboratoriais e projetos colaborativos.

- O que significa uma ferramenta ter acesso via API?
 - Significa que pode ser integrada automaticamente com outros sistemas, permitindo troca de dados e automatização de processos.

- De que forma posso otimizar a minha revisão de literatura para além das tradicionais pesquisas por palavras-chave?
 - Recomenda-se o uso de ferramentas de mapeamento de literatura, como o Litmaps ou o ResearchRabbit. Estas plataformas utilizam IA e gráficos de conhecimento para visualizar cadeias de citação, permitindo descobrir artigos relevantes que podem não aparecer em buscas convencionais (como as do Google Scholar).

- Quais são as ferramentas padrão para a limpeza e tratamento de grandes volumes de dados de investigação?
 - O OpenRefine é a ferramenta de referência para a limpeza e normalização de dados. No entanto, para fluxos de trabalho mais complexos ou volumes massivos de informação, as opções mais robustas são o KNIME ou o Apache Spark.

- Onde devo alojar o software e as publicações do meu projeto para garantir o acesso aberto?
 - Para software e conjuntos de dados volumosos (até 50 GB), o Zenodo é o repositório internacional mais indicado. No contexto nacional português, as publicações devem ser depositadas no RCAAP (Repositório Científico de Acesso Aberto de Portugal) para assegurar a visibilidade e o cumprimento das normas institucionais.

- É seguro utilizar ferramentas de IA (p.ex. ChatPDF) para analisar documentos de investigação?
 - Sim, para documentos já públicos, mas com reservas: a informação extraída deve ser cruzada com outras fontes devido ao risco de alucinações da IA. Dados confidenciais ou não publicados nunca devem ser inseridos em ferramentas de IA generativa não aprovadas pela instituição.

- Como garantir a preservação de software a muito longo prazo (séculos)?
 - Além de repositórios como o GitHub ou Zenodo, recomenda-se o uso do Software Heritage, que funciona como um arquivo global para o código-fonte, visando a sua durabilidade histórica.

- Quais as vantagens de usar ferramentas de mapeamento visual (p.ex. Litmaps ou ResearchRabbit) em relação ao Google Scholar?
 - Estas ferramentas utilizam **gráficos de conhecimento** e IA para identificar cadeias de citação e "saltos" semânticos. Permitem visualizar a evolução de um tópico, identificar autores centrais e detetar artigos relevantes que não contêm as palavras-chave exatas da sua pesquisa.

(Fonte:<https://effortlessacademic.com/litmaps-vs-researchrabbit-vs-connected-papers-the-best-literature-review-tool-in-2025/>)

- Qual a ferramenta recomendada para a limpeza e normalização de grandes volumes de dados desordenados?
 - O **OpenRefine** continua a ser o padrão para curadoria de dados em larga escala (limpeza de nomes, fusão de entidades via algoritmos de *clustering*). Para fluxos de trabalho visuais mais complexos, o **KNIME** ou **Apache Spark** são preferenciais em contextos de ciência de dados avançada.

(Fonte:<https://openrefine.org/>)

Ética e Transparência

- Porque é que a transparência é essencial para a confiança na ciência?
 - A transparência permite a verificação independente, reduz a fraude e aumenta a credibilidade da ciência junto da sociedade, especialmente em contextos de financiamento público.
- O que é o Pré-registo?
 - A publicação da metodologia e do plano de análise antes de recolher os dados, prevenindo viés e fraude (como p.ex. p-hacking ou HARKing).
- O que é valorizado pelo DORA?
 - A avaliação da investigação deve focar-se na qualidade, impacto e na diversidade de outputs (não apenas publicações de artigos em revistas).
- Posso usar dados de redes sociais?
 - Sim, mas com transparência e o RGPD exige a avaliação da expectativa razoável de privacidade dos indivíduos, ou uso de dados anonimizados de forma robusta.
- O que é o DAS (Data Availability Statement)?
 - O DAS é uma declaração obrigatória que deve constar em todas as publicações científicas.
- Qual a finalidade do DAS?
 - Informar explicitamente os leitores sobre a localização dos dados de suporte (PID) e as condições de acesso e reutilização desses dados. O DAS é crucial para cumprir os princípios da Localização (F) e Acessibilidade (A) do FAIR.
- Posso indicar no DAS para contactar o autor para obter os dados?
 - Não. Isso é considerado uma falha de acessibilidade (Accessibility issue) e deve ser evitado; o DAS deve fornecer um PID ou URL persistente.
- DAS é necessário mesmo se os dados estiverem fechados?
 - Sim, deve descrever a localização (PID) e as condições de acesso restrito (managed access procedure), garantindo a transparência, preferencialmente através de um DOI para os metadados.

- Qual o benefício de integrar Práticas Recomendadas de CA em propostas Horizon Europe?
 - Integrar práticas recomendadas, como Open Peer Review ou Early Sharing (pré-registos, preprints), resultará numa pontuação de avaliação mais alta no aspecto da Ciência Aberta (Metodologia).

- O que é Co-criação (Co-creation) no Horizonte Europa?
 - É a abordagem colaborativa onde stakeholders (cidadãos, indústria, decisores políticos) moldam ativamente o projeto, desde o design à implementação, acelerando o impacto.

- Que ferramentas ajudam a escrever um DAS?
 - Embora existam poucas ferramentas de "gerador automático", ferramentas de escrita académica baseadas em IA (ResearchPal) podem redigir um DAS se o contexto for fornecido. Grandes editoras (Elsevier, Springer Nature, Taylor & Francis) possuem "wizards" integrados nos seus sistemas de submissão que geram o texto automaticamente. A geração de textos por IA não isenta os utilizadores da responsabilidade, devendo verificar os resultados.

- Qual a forma mais eficaz de criar um DAS?
 - Utilizar os templates mais aceites universalmente e adaptá-los à situação específica.
 - Modelos Padrão (Exemplos Aalto):
 - Cenário A (Abertamente Disponível): "The data that support the findings of this study are openly available in [Nome do Repositório] at [URL/DOI], reference number [número]."
 - Cenário B (Mediante Solicitação): "The data that support the findings of this study are available from the corresponding author, [Iniciais do Autor], upon reasonable request."
 - Guias Oficiais das Editoras:
 - Se souber a revista de destino, use o guia específico, como p.ex. o Springer Nature DAS Helper (lista exaustiva de frases) ou o Taylor & Francis Author Services (tabela de cenários vs. declarações).

- Resultados negativos devem ser partilhados?
 - Sempre que possível, sim. A partilha de resultados negativos contribui para a transparência e evita duplicação de esforços.

- Como evitar revistas predatórias?
 - Utilizando listas e ferramentas de verificação reconhecidas e avaliando critérios como revisão por pares, transparência editorial e indexação.
- O que fazer se identificar um erro após a publicação?
 - Deve contactar a revista ou editora e seguir os procedimentos formais de correção ou retratação
- Como é que as instituições e editoras estão a combater as "Paper Mills" (fábricas de artigos fraudulentos)?
 - Atualmente, utiliza-se o STM Integrity Hub, uma plataforma que deteta fraudes sistémicas. Este sistema é capaz de identificar sinais de alerta, como "frases torturadas" (traduções ou paráfrases bizarras geradas por algoritmos) e padrões de submissão suspeitos que indicam produção em massa de artigos falsos.

(Fonte: <https://stm-assoc.org/integrity-hub/>)

- Existem ferramentas automáticas para verificar a integridade de imagens científicas antes da submissão?
 - Sim, ferramentas de visão computacional como **ImageTwin** e **Proofig AI** detetam duplicações de bandas (Western Blot), rotações, cortes e até manipulações de contraste impercetíveis ao olho humano, comparando a imagem com bases de dados de milhões de figuras já publicadas.

(Fonte: <https://imagetwin.ai/solutions/institutions>)

- O que é um Registered Report e qual a sua vantagem para a integridade científica?
 - Um Registered Report é um modelo de publicação onde o design do estudo e a metodologia são revistos por pares antes da recolha de dados. Se o design for aprovado, a revista garante a publicação independentemente dos resultados, o que combate o viés de publicação (a tendência de publicar apenas resultados positivos).
- Como garantir que o código e o ambiente de análise de um estudo permanecem reprodutíveis no futuro?
 - A solução passa pela utilização de Cápsulas de Computação, em plataformas como o Code Ocean. Estas cápsulas empacotam o código, os dados e todo o ambiente de software (usando tecnologia como o Docker), garantindo que qualquer investigador consiga correr a análise exatamente da mesma forma, perpetuamente.

- Qual a vantagem dos Registered Reports para a integridade científica?
 - Garantem que a decisão de publicar se baseia na importância da pergunta de investigação e no rigor do método, independentemente de os resultados serem positivos, negativos ou nulos.

- O que são Registered Reports e qual a sua vantagem?
 - É um modelo onde o protocolo de investigação é avaliado antes da execução. A vantagem é garantir a publicação baseada no rigor do método e não apenas em resultados positivos ou espetaculares, combatendo o viés científico.

- É obrigatório ter planos de contingência para fugas de dados?
 - Sim. Em sistemas de acesso controlado com dados sensíveis, o PGD deve prever procedimentos em caso de ataque, p.ex. quem desliga o sistema e o dever de informar obrigatoriamente os participantes afetados.

- Como obter consentimento de comunidades iletradas ou vulneráveis?
 - A comunicação deve ser adaptada usando suportes visuais (p.ex. vídeos, imagens ou slides) ou recorrendo a um mediador/tradutor da própria comunidade que explique a natureza e destino dos dados, garantindo que o consentimento seja explícito e reversível.

Reprodutibilidade

- Qual a diferença entre Reprodutibilidade e Replicabilidade?
 - Reprodutibilidade é verificação (mesmos dados/código): reproduzir os mesmos resultados; Replicabilidade é validação (novo dado/mesmo método): replicar o método com dados diferentes.
- A reprodutibilidade é apenas técnica?
 - Não. Para além da reprodutibilidade computacional, é essencial a reprodutibilidade conceptual, baseada em metodologias claras, documentação adequada e descrição completa do contexto científico.
- Qual é o principal benefício do Jupyter/Binder para a reprodutibilidade?
 - Permite que qualquer pessoa execute a análise original no ambiente exato usado pelo investigador, resolvendo o problema das configurações e dependências de software (dependências).
- O que deve conter um README file de dados?
 - Deve incluir o contexto da investigação, a estrutura dos dados, um dicionário de dados (codebook), a licença e o contacto para dúvidas.
- O que é Provenance (Proveniência)?
 - É o registo histórico detalhado de todas as transformações, versões e agentes (pessoas/ferramentas) que processaram os dados, essencial para a integridade.
- Os notebooks devem ser partilhados?
 - Sim, são outputs valiosos para a validação e reutilização, e devem ser partilhados em repositórios, tal como outros outputs de investigação.
- Onde partilhar workflows?
 - P.ex. no WorkflowHub (repositório registado) para que os fluxos de trabalho computacionais sejam produtos académicos formais e possam ser citados.
- Como garantir que o código funciona no futuro?
 - Usando ferramentas que encapsulam o ambiente computacional, como Jupyter Notebooks e Binder.

- O que é Interoperabilidade Legal?
 - A capacidade de combinar e reutilizar dados de diferentes jurisdições sem violar leis de PI ou de privacidade, exigindo clareza e compatibilidade de licenças (CODATA-RDA).

- A reprodutibilidade exige sempre partilha de dados?
 - Nem sempre. Em casos de dados sensíveis, a reprodutibilidade pode ser assegurada através de descrições metodológicas detalhadas e acesso controlado.

- É obrigatório disponibilizar o código-fonte?
 - Sempre que o código seja essencial para os resultados, a sua disponibilização é fortemente recomendada.

- Qual o papel dos containers na reprodutibilidade?
 - Os containers permitem recriar ambientes computacionais consistentes, facilitando a reprodução das análises.

- Qual a diferença prática entre um ficheiro README e os metadados formais?
 - Os metadados são etiquetas descritivas muitas vezes legíveis por máquina; o README é um suporte explicativo humano que detalha o contexto, critérios de exclusão de dados (ruído) e transformações efetuadas, essenciais para quem tenta reproduzir o estudo.

- Como assegurar a precisão em observações humanas subjetivas (p.ex. identificar animais pelas suas pegadas)?
 - Deve-se documentar rigorosamente os elementos de treino e as orientações sistemáticas (convenções de anotação) seguidas pelos observadores para que, mesmo com a rotatividade da equipa, os critérios de registo permaneçam consistentes.

- O que é um "Registered Report" e como ajuda na reprodutibilidade?
 - É um modelo de publicação em duas fases. Na primeira fase, o design do estudo e o plano de análise são revistos por pares *antes* dos dados serem recolhidos. Se aprovado, a revista garante a publicação independentemente dos resultados serem positivos ou negativos. Isto combate o "viés de publicação" e incentiva métodos robustos.

Fonte: [Center for Open Science \(COS\) - Registered Reports](#)

- O que são "Cápsulas de Computação" (Compute Capsules) e como garantem a reprodutibilidade perpétua?
 - Plataformas como o Code Ocean empacotam o código, os dados, o ambiente (Docker) e as dependências num recipiente imutável. Isto garante que qualquer investigador possa executar a análise e obter os mesmos resultados daqui a 10 anos, independentemente das atualizações de hardware ou software.

(Fonte:<https://codeocean.com/>)

- Como o Renku facilita o rastreio da proveniência de dados e código?
 - O Renku automatiza o registo de todos os passos da análise via um Gráfico de Conhecimento. Cada ficheiro de saída está ligado ao código e aos dados de entrada exatos que o geraram, permitindo a reexecução total do *workflow* e a auditoria de resultados.

(Fonte:<https://renkulab.io/>)

Princípios de Gestão de Dados de Investigação e DMPs

- O PGD é final e definitivo?
 - Não, o PGD deve ser um documento dinâmico e atualizado regularmente sempre que ocorrerem mudanças significativas no projeto.
- Com que frequência devo atualizar o meu Plano de Gestão de Dados (PGD)?
 - O PGD é um documento vivo. Deve ser atualizado em três momentos chave: 1) No início do projeto (proposta); 2) A meio do projeto (ou sempre que houver alterações significativas na metodologia ou volume de dados); 3) No final do projeto, para refletir o estado final dos dados preservados.

Link: [FCT - Guia de Gestão de Dados e PGD](#)

- Quais são os 6 requisitos centrais para um PGD?
 - 1. Descrição dos dados;
 - 2. Documentação e qualidade;
 - 3. Armazenamento e segurança;
 - 4. Requisitos legais/éticos;
 - 5. Partilha e preservação;
 - 6. Responsabilidades e recursos.
- O PDG é obrigatório no Horizonte Europa?
 - Sim. O PGD é um deliverable obrigatório (primeira versão até ao Mês 6 do projeto) registado como um deliverable público não restrito.
- Quais são os principais elementos de conteúdo do PGD?
 - Descrição do dataset (tipo, tamanho, origem), standards e metadados, identificadores persistentes, metodologia de curadoria e preservação, metodologia de partilha e custos.
- O PGD deve incluir custos?
 - Sim. Deve justificar os recursos (financeiros e tempo) dedicados a tornar os dados FAIR (curadoria, armazenamento), pois são elegíveis em programas de financiamento, e relevantes para a preservação no longo prazo.

- Os custos de RDM são elegíveis?
 - Sim, o tempo de pessoal e os custos de armazenamento e curadoria para tornar os dados FAIR são elegíveis para financiamento (p.ex. Horizonte Europa).

- Qual o objetivo da Rede de Data Stewards?
 - Reunir profissionais com interesses e desafios comuns na GDI, promovendo um espaço de partilha e capacitação.

- O Plano de Gestão de Dados deve ser público?
 - Sempre que possível, sim, podendo ser disponibilizado com informação sensível removida ou resumida.

- Com que frequência o PGD deve ser atualizado?
 - Deve ser revisto sempre que ocorram alterações significativas no projeto ou nos dados.

- Quem é responsável pelo PGD?

A responsabilidade é do investigador principal, podendo ser partilhada com data stewards ou equipas de apoio.

- Posso automatizar o PGD?
 - Sim, através de maDMP APIs, que permitem ao PGD "ouvir" atualizações de outros sistemas (p.ex. receber um DOI automaticamente). O DAMAP ajuda a pré-preencher informações administrativas integrando-se com sistemas da universidade.

(Fonte: <https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard>)

- Como o DAMAP facilita a integração do PGD com os sistemas da universidade?
 - O DAMAP é uma ferramenta de maDMP que se integra com sistemas CRIS (Research Info) e bases de dados de RH. Ele pré-preenche as informações administrativas e financeiras do PGD, permitindo que o investigador se foque apenas nas especificidades técnicas dos dados.

(Fonte: <https://forschungsdaten.at/en/damap/>)

- Como classificar dados de investigação?
 - Em públicos, internos, confidenciais e sensíveis.

- Dados sensíveis podem ser FAIR?
 - Sim, através de metadados ricos e acesso controlado.

- O que é governação de dados em investigação?
 - Conjunto de políticas, papéis e responsabilidades que asseguram uma gestão legal, ética e sustentável dos dados.

- Qual a diferença entre Data Owner, Data Controller e Data Processor?
 - Data Owner: entidade com responsabilidade institucional.
 - Data Controller: define finalidades e meios do tratamento (RGPD).
 - Data Processor: trata dados em nome do controller.

- Qual o papel dos serviços institucionais?
 - Bibliotecas, IT, jurídico e data stewards apoiam curadoria, licenciamento, conformidade e preservação.

- O PGD pode ser usado em auditorias?
 - Sim. O PGD é frequentemente utilizado em auditorias intermédias e finais para verificar se a gestão de dados declarada no PGD corresponde à prática efetiva do projeto.

- Como posso garantir que o código da minha investigação pode ser executado por outros?
 - Utilizando plataformas como o Code Ocean ou o MyBinder, que encapsulam o código e o ambiente de software, permitindo a execução imediata sem necessidade de instalações complexas.

- O uso da ferramenta Argos é exclusivo para projetos com financiamento?
 - Não. É possível criar planos para projetos sem financiamento usando templates genéricos ou institucionais. A obrigatoriedade de identificar o financiamento apenas surge ao selecionar um blueprint específico de um financiador (p.ex. FCT ou Horizonte Europa).

- Deve-se criar um PGD por projeto ou um plano global para o grupo de investigação?
 - É obrigatório um PGD por projeto em contextos financiados. Contudo, é uma boa prática manter templates de datasets ao nível do grupo de investigação no Argos, que podem ser clonados e reutilizados para garantir eficiência e continuidade.

Utilização de Inteligência Artificial (IA) em Dados de Investigação

- A IA pode ser autora de um artigo?
 - Não. A autoria não pode ser atribuída à IA, e o investigador é sempre o responsável final por qualquer output gerado.
- A IA pode ser coautora?
 - Não. A maioria das editoras científicas e organizações de ética (como o COPE) estabeleceram que ferramentas de IA não podem ser autores porque não podem assumir responsabilidade legal ou ética pelo conteúdo. O uso de IA deve ser declarado explicitamente na secção de Metodologia ou Agradecimentos.

Fonte: [COPE - Authorship and AI Tools](#)

- O que devo declarar sobre o uso de IA?
 - Detalhar de forma transparente o uso substancial de ferramentas de IA generativa, medidas de fiabilidade e verificação, e registar uso para reprodutibilidade futura.
 - Por exemplo, descrever o modelo/versão da IA usada, a origem dos dados de treino (para mitigar o viés) e as medidas de segurança
- Quais os principais riscos da IA?
 - Fuga de informação confidencial/pessoal e geração de alucinações p.ex. conteúdos e citações falsas.
- Existe alguma isenção legal para a investigação?
 - Sim, o AI Act geralmente não se aplica a sistemas de IA usados exclusivamente para fins de Investigação e Desenvolvimento (R&D) em laboratório.
- Posso usar IA para analisar dados sensíveis?
 - Apenas se forem cumpridos os requisitos legais, éticos e de segurança, incluindo avaliação de risco e controlo de acesso.
- Devo declarar o uso de IA em propostas e publicações?
 - Sim. A transparência no uso de IA é uma boa prática recomendada por editores e organismos de ética.

- A IA pode comprometer a reprodutibilidade?
 - Pode, se não forem documentadas as ferramentas, versões e parâmetros utilizados. A documentação adequada mitiga este risco.

Qual é a principal diferença entre utilizar o ChatGPT e ferramentas de IA académica como o Consensus ou o Elicit na investigação?

 - Diferente do ChatGPT genérico, estas ferramentas utilizam **RAG (Retrieval-Augmented Generation)** ligado a bases de dados científicas indexadas (como Semantic Scholar ou OpenAlex). Todas as respostas são obrigatoriamente ancoradas em artigos reais com citações verificáveis e clicáveis.

(Fonte: <https://consensus.app/>)

- Como posso garantir a integridade ética ao utilizar Inteligência Artificial num projeto de investigação?
 - Para projetos que envolvem IA, é fundamental realizar uma **Avaliação de Impacto Ético (EIA)**. Este processo serve para identificar e mitigar riscos antes da implementação da tecnologia. A recomendação atual é a utilização do **AI Ethics Toolkit da UNESCO**, que oferece uma estrutura prática para avaliar dois pilares críticos:
 - Discriminação:** Garantir que os algoritmos não perpetuam preconceitos ou exclusões.
 - Transparência:** Assegurar que os processos de decisão da IA sejam compreensíveis e auditáveis pelos investigadores e pela sociedade.
- Datasets gerados ou enriquecidos por IA podem ser depositados?
 - Sim, desde que a metodologia, ferramentas e versões estejam claramente documentadas.
- Como posso garantir que os dados que uso para treinar IA são éticos?
 - Deve verificar se os dados foram recolhidos com consentimento informado explícito para fins de treino de IA, se estão livres de enviesamentos (*biases*) históricos conhecidos e se a utilização respeita a propriedade intelectual. Recomenda-se a documentação através de "Datasheets for Datasets".

Link: [Microsoft Research - Datasheets for Datasets Guidance](#)

- Como se deve realizar uma Avaliação de Impacto Ético (EIA) para projetos que utilizam IA?
 - Recomenda-se a utilização do AI Ethics Toolkit da UNESCO. Ele fornece o *Ethical Impact Assessment (EIA)*, um roteiro estruturado para avaliar riscos

de discriminação, transparência e responsabilidade humana no uso de algoritmos, alinhado com o AI Act da UE.

Fonte: (<https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>)

- Como garantir a reprodutibilidade e transparência ao usar IA generativa na escrita ou análise?
 - Não basta citar a ferramenta (p.ex. ChatGPT). Para garantir a reprodutibilidade, deve-se documentar e, se possível, publicar os prompts utilizados (Prompt Engineering). As boas práticas incluem guardar os registos das interações e verificar se a IA foi usada apenas como ferramenta linguística ou se gerou conteúdo substantivo, devendo este último ser explicitamente demarcado para evitar alucinações ou plágio involuntário.

- Todos os projetos com IA exigem Avaliação de Impacto Ético?
 - Nem todos, mas sempre que a IA possa afetar direitos fundamentais, introduzir enviesamentos ou tratar dados sensíveis, a Avaliação de Impacto Ético é considerada boa prática e, em alguns casos, obrigatória.

- Qual a utilidade de registar os "inputs" (prompts) perante a natureza estocástica da IA?
 - Segue o princípio da completude na documentação e serve para autorreferência interna ou comparação futura entre modelos.

- O Microsoft Copilot é semelhante ao ChatGPT?
 - A tecnologia de base é idêntica, sendo um sistema hospedado nos EUA. A diferença prática é estar pré-instalado no Windows, em vez de exigir acesso via browser.

- Como garantir a ética nos dados usados para treinar IA?
 - Verificar se houve consentimento informado explícito para treino de algoritmos, mitigar enviesamentos históricos e documentar via "Datasheets for Datasets"

Links e Recursos Essenciais para Ciência Aberta e GDI

Categoria	Recurso	Link
Financiadores e Políticas Institucionais	FCT Política de Acesso Aberto (Regras e Regulamentos)	https://www.fct.pt/sobre/estudos-e-planeamento-estrategico/politicas-de-ciencia-aberta/acesso-aberto-a-publicacoes-cientificas/
	UNESCO	
	Recomendação sobre Ciência Aberta	https://unesdoc.unesco.org/ark:/48223/pf0000379949
	EU Strategic Principles (Política de Ciência Aberta da UE)	https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science_en#the-eus-open-science-policy
	Barcelona Declaration on Open Research Information (Declaração de Princípios)	https://barcelona-declaration.org/downloads/BarcelonaDeclaration.pdf
	Lei da Ciência (Decreto-Lei n.º 63/2019)	https://files.dre.pt/1s/2019/05/09400/0246602475.pdf
	Resolução do Conselho de Ministros n.º 21/2016 (Política Nacional de CA)	https://diariodarepublica.pt/dr/detalhe/resolucao-conselho-ministros/21-2016-74094659
Infraestruturas e Repositórios	Princípios de Infraestrutura Académica Aberta (POSI)	https://openscholarlyinfrastructure.org/
	European Open Science Cloud (EOSC)	https://open-science-cloud.ec.europa.eu/
	Re.Data — Catálogo e Recursos	https://redata.pt/

	<p>Específicos (Portugal)</p> <p>Registry of Research Data Repositories (Re3data)</p> <p>Open Science Framework (OSF)</p> <p>Software Heritage</p> <p>Open Research Europe (ORE) (Plataforma de Publicação UE)</p> <p>Diretório de Repositórios da FCCN (RCAAP)</p> <p>Registo de Políticas de Auto-arquivo (SHERPA ROMEO)</p> <p>Catalogue — Data Sources (Catálogo de Dados Aalto)</p> <p>CoreTrustSeal</p>	<p>https://www.re3data.org/</p> <p>https://osf.io/</p> <p>https://www.softwareheritage.org/</p> <p>https://open-research-europe.ec.europa.eu/</p> <p>https://www.rcaap.pt/</p> <p>https://openpolicyfinder.jisc.ac.uk/</p> <p>https://datahub.aalto.fi/en/data-sources</p> <p>https://www.coretrustseal.org/</p>
Padrões de Dados e Metadados	<p>Fairsharing.org (Catálogo de Standards, Bases de Dados e Políticas)</p> <p>Research Data Alliance (RDA)</p> <p>Diretório de Metadados RDA</p> <p>Ferramenta FAIR- by-Design (Orientação para Materiais Didáticos)</p>	<p>https://fairsharing.org/</p> <p>https://www.rd-alliance.org/</p> <p>http://rd-alliance.github.io/metadata-directory/</p> <p>https://www.openaire.eu/how-to-make-your-data-fair</p>

	FAIR-Aware (ferramenta de autoavaliação FAIR)	https://fairaware.dans.knaw.nl/
	F-UJI FAIR Assessment Tool	https://www.f-uji.net/
	Diretrizes para Curadoria de Metadados (Re.Data)	https://redata.pt/curadoria-de-metadados-em-repositorios-de-dados-de-investigacao/
Ferramentas de Planeamento (PGD)	ARGOS (Ferramenta Open Source para maDMPs)	https://argos.openaire.eu/
	Guia de Utilizador ARGOS (FCT Modelo)	https://argos.openaire.eu/user-guide
	DMPonline (Plataforma PGD)	https://dmponline.dcc.ac.uk/
	OpenAIRE Costing Tool (https://www.openaire.eu/costing-tool
Licenciamento e Citação	Creative Commons License Chooser (Ferramenta de Seleção de Licenças)	https://creativecommons.org/chooser/
	Opção alternativa de seleção de licença	https://chooselicense.com/
	Diretório de Revistas de Acesso Aberto (DOAJ)	https://doaj.org/
	Journal Checker Tool (JCT) (Conformidade Plans)	https://journalcheckertool.org/
Guias e Formação	OpenAIRE Guide (Guias de Conformidade HE)	https://www.openaire.eu/guides

	FOSTER Open Science	https://www.fosterscience.eu/
	FIT4RI project guidelines	https://fit4ri.eu/guidelines/
	GitHub Guides	https://docs.github.com/
	PATHOS Open Science Handbook (Indicadores de Impacto)	https://handbook.pathos-project.eu/
	PATHOS OS Stories Registry	https://pathos-project.eu/os-stories-registry
	Documentação da Plataforma EOSC File Sync & Share	https://open-science-cloud.ec.europa.eu/services
	Open Science Policy Checklist (Checklist para RPOs)	https://www.openaire.eu/open-science-policy-checklist-for-research-performing-organisations
Outros Recursos	Relatório de Avaliação em HE (Zenodo)	https://doi.org/10.5281/zenodo.7324363
	ALLEA – European Code of Conduct for Research Integrity	https://allea.org/code-of-conduct/
	European Citizen Science Association (ECSA)	https://ecsa.citizen-science.net/