



Política de Dados Abertos



Sumário

Política de Dados Abertos	1
Glossário	3
Introdução e Motivação Estratégica	4
Definição de Dados de Investigação	5
Âmbito e Princípios Orientadores	5
Repositório de dados de investigação do INESC TEC.....	7
Publicação de Dados de Investigação e Compromissos Institucionais.....	7
Princípios Gerais.....	7
Orientação para Investigadores.....	8
Papéis e Responsabilidades.....	10
Conclusão	11

Glossário

Plano de Gestão de Dados (DMP)

Documento que descreve a forma como os dados de investigação serão geridos ao longo do seu ciclo de vida, incluindo aspetos como a recolha, organização, armazenamento, condições de acesso, partilha, publicação, preservação e responsabilidades associadas.

Data Stewardship

Conjunto de práticas, responsabilidades e processos que suportam a gestão, documentação, preservação, partilha e reutilização eficazes dos dados de investigação ao longo do seu ciclo de vida, em conformidade com os princípios FAIR e com os requisitos legais, éticos e contratuais aplicáveis.

Dataset (Conjunto de dados)

Unidade de dados de investigação constituída por um ou mais ficheiros, organizada e documentada de forma a permitir a sua descoberta, validação, interpretação, reutilização ou integração em atividades de investigação posteriores.

Identificador Persistente de Objeto Digital (DOI)

Tipo amplamente utilizado de identificador persistente atribuído a *datasets* e outros resultados de investigação para permitir a sua identificação única, citação, descoberta e acesso a longo prazo.

Princípios FAIR

Conjunto de princípios orientadores destinados a tornar os dados de investigação Encontráveis (*Findable*), Acessíveis (*Accessible*), Interoperáveis (*Interoperable*) e Reutilizáveis (*Reusable*), aumentando a sua visibilidade, acessibilidade e potencial de reutilização.

Metadados

Informação estruturada que descreve um dataset e facilita a sua descoberta, interpretação, gestão, citação e reutilização.

Dados Abertos

Dados de investigação disponibilizados para acesso, utilização e reutilização com o mínimo de restrições possível, sujeitos às limitações legais, éticas, contratuais ou de propriedade intelectual aplicáveis.

Dados de investigação

Registos factuais e informações gerados, recolhidos, observados ou processados no âmbito de atividades de investigação. Os dados de investigação podem ser quantitativos ou qualitativos, brutos ou derivados, estruturados ou não estruturados, e incluem todos os materiais necessários para suportar a validação, interpretação e reutilização dos resultados da investigação.

Introdução e Motivação Estratégica

A crescente ênfase na transparência, na reprodutibilidade e na responsabilidade pública tem colocado os dados abertos no centro das políticas científicas internacionais. Na Europa, iniciativas como a European Open Science Cloud (EOSC), os princípios FAIR (*Findable, Accessible, Interoperable, Reusable*) e as orientações da Comissão Europeia refletem um consenso claro: os dados resultantes de investigação financiada com fundos públicos devem ser tratados como um bem público.

Estes desenvolvimentos enquadram-se também em novas abordagens à avaliação científica. Através da Coalition for Advancing Research Assessment (CoARA), os dados de investigação passam a ser reconhecidos como um *output* científico relevante, com impacto na avaliação e progressão das carreiras. Em paralelo, práticas de dados abertos alinhadas com os princípios FAIR contribuem para os Objetivos de Desenvolvimento Sustentável (ODS), reforçando o valor social da ciência e promovendo soluções baseadas em evidência para desafios globais.

Para o INESC TEC, a adoção destes princípios vai além da conformidade. Representa uma oportunidade estratégica para reforçar a sua reputação científica, promover colaborações internacionais e aumentar o impacto societal da sua investigação.

Este compromisso é reforçado pelo reconhecimento do INESC TEC como um dos centros nacionais na área da Gestão de Dados de Investigação. Este papel reforça a consolidação de práticas robustas e sustentáveis de Data Stewardship em toda a instituição.

A gestão adequada, preservação e partilha de dados - *tão abertos quanto possível e tão restritos quanto necessário* - são hoje fundamentais. Estas práticas sustentam a integridade científica e promovem a colaboração e a inovação.

Neste contexto, o INESC TEC reconhece os dados de investigação como um resultado científico relevante e, como princípio geral, um bem público, reconhecendo simultaneamente que restrições legítimas de natureza legal, ética, contratual ou relacionadas com a propriedade intelectual podem justificar a limitação do seu acesso.

A presente Política de Dados Abertos define a abordagem institucional à gestão e disseminação de dados de investigação, estabelecendo orientações alinhadas com os princípios FAIR para garantir a sua partilha responsável, acessibilidade a longo prazo e reutilização. Constitui, assim, um instrumento essencial de suporte à estratégia de Ciência Aberta do INESC TEC, conforme definida no Plano Estratégico 2023–2030.

Enquanto organização de investigação e tecnologia, o INESC TEC encara também os dados de investigação como um ativo estratégico institucional. A sua gestão adequada reforça o valor, a visibilidade e a utilização a longo prazo dos resultados científicos, contribuindo não apenas para a conformidade e abertura, mas também para maximizar o impacto científico, societal e organizacional dos dados produzidos.

Com a integração dos princípios FAIR nos seus processos, o INESC TEC pretende:

- Reforçar a visibilidade, acessibilidade e impacto dos resultados de investigação;
- Garantir a integridade, rastreabilidade e reprodutibilidade dos dados;

- Alinhar-se com requisitos europeus e internacionais de partilha responsável de dados;
- Promover uma cultura institucional de abertura e colaboração;
- Aumentar o reconhecimento e a competitividade no panorama científico internacional.

Definição de Dados de Investigação

Para efeitos da presente política, consideram-se dados de investigação todos os dados gerados, recolhidos, observados ou tratados no âmbito de atividades de investigação. De acordo com definições internacionais, nomeadamente da OCDE e da Comissão Europeia, estes dados podem assumir diferentes formas: quantitativas ou qualitativas, brutas ou derivadas, estruturadas ou não estruturadas. Incluem, por exemplo, medições experimentais, respostas a inquéritos, entrevistas, imagens, materiais audiovisuais, modelos ou resultados de simulações.

No âmbito desta política, um *dataset* corresponde a uma unidade organizada de dados de investigação, constituída por um ou mais ficheiros, estruturada e documentada de forma a permitir a sua descoberta, validação, reutilização ou integração em atividades de investigação posteriores. Os *datasets* são considerados resultados científicos autónomos, contribuindo para a transparência, reprodutibilidade e colaboração.

Esta definição abrange também outputs derivados que funcionam como *datasets*, como conjuntos de dados de referência (*benchmark*), dados de treino ou avaliação (incluindo corpora em áreas como o processamento de linguagem natural), materiais anotados ou resultados estruturados de simulações. O elemento comum a estes *outputs* é o seu potencial de reutilização por terceiros, seja para validação de resultados, comparação de resultados, replicação de experiências ou realização de novas análises.

Para garantir a sua utilização efetiva, os *datasets* devem ser acompanhados de documentação adequada, metadados e, sempre que aplicável, materiais auxiliares, incluindo scripts ou código, que permitam a sua validação, interpretação e reutilização.

Âmbito e Princípios Orientadores

A presente política define a abordagem institucional do INESC TEC à gestão responsável, preservação e partilha de dados de investigação, em alinhamento com os padrões europeus e com os princípios FAIR. Aplica-se a:

- Dados de investigação produzidos, recolhidos ou significativamente tratados por investigadores do INESC TEC no âmbito de atividades científicas, independentemente de serem financiadas externamente ou apoiadas internamente;
- Dados que suportam publicações científicas ou que constituam, por si só, resultados de investigação;
- Dados primários e secundários, estruturados ou não estruturados, quantitativos ou qualitativos;

- Dados que não possam ser disponibilizados em acesso aberto, desde que as restrições de acesso se baseiem em fundamentos legítimos de natureza ética, legal ou contratual.

Esta política não substitui a legislação aplicável nem os requisitos específicos definidos por entidades financiadoras. Pelo contrário, procura alinhar as práticas internas do INESC TEC com essas obrigações externas, proporcionando aos investigadores um enquadramento claro que assegure conformidade, qualidade e impacto.

A implementação desta política assenta nos princípios *FAIR*, que constituem a base das práticas de gestão e partilha de dados no INESC TEC:

Findable (Localizáveis): os *datasets* devem possuir identificadores persistentes, incluir metadados ricos e normalizados e estar registados em repositórios pesquisáveis;

Accessible (Acessíveis): os *datasets* devem ser recuperáveis através de protocolos abertos e normalizados, com condições de acesso claramente definidas;

Interoperable (Interoperáveis): os *datasets* devem ser preparados e partilhados em formatos estruturados, com metadados que sigam normas reconhecidas;

Reusable (Reutilizáveis): os *datasets* devem estar bem documentados, incluir informação clara sobre licenciamento e cumprir as práticas da comunidade científica relevante.

Enquanto instituição de investigação, o INESC TEC compromete-se a:

- Assegurar a preservação a longo prazo e a reutilização responsável dos dados de investigação;
- Promover o acesso aberto aos *datasets*, sempre que possível em conformidade com FAIR;
- Incentivar boas práticas de gestão de dados e o desenvolvimento de competências na comunidade científica;
- Fomentar uma cultura de ciência aberta e colaborativa, suportada por orientações claras e infraestruturas institucionais adequadas.

A implementação eficaz desta política requer mecanismos institucionais que permitam aos investigadores gerir e partilhar dados ao longo de todo o ciclo de vida da investigação. A secção seguinte apresenta as plataformas e serviços disponíveis para apoiar estas práticas, bem como orientações relativas à documentação, licenciamento, utilização de repositórios e acesso a longo prazo.

Para mais informações ou apoio individualizado, os investigadores podem contactar o Serviço de Apoio à Gestão.

Repositório de dados de investigação do INESC TEC

O INESC TEC disponibiliza um Repositório Institucional de Dados de Investigação (*Research Data Management - RDM*) com o objetivo de apoiar a implementação desta política e promover a partilha responsável de dados de investigação.

O [RDM INESC TEC](#) reflete o compromisso institucional com práticas de dados abertos e alinhadas com os princípios FAIR e inclui *datasets* produzidos por investigadores e colaboradores do INESC TEC em diversas áreas científicas. O repositório destina-se principalmente à publicação, preservação e descoberta de *datasets* para partilha e reutilização, não tendo como objetivo substituir infraestruturas especializadas concebidas para ambientes de dados dinâmicos, operacionais ou de grande escala.

O repositório baseia-se na plataforma de código aberto *CKAN* e permite a atribuição de identificadores persistentes (*DOI - Digital Object Identifier*), através do serviço *DataCite Fabrica*. Assegura também a preservação a longo prazo, através de cópias de segurança diárias, e inclui informação de citação como parte integrante dos metadados.

Este repositório constitui igualmente um elemento central no apoio à implementação de Planos de Gestão de Dados (*Data Management Plans - DMPs*), habitualmente elaborados ao nível dos projetos para documentar a gestão dos dados de investigação ao longo do seu ciclo de vida. Encontra-se registado no [re3data.org](#) e indexado no [FAIRsharing.org](#), o que garante a sua visibilidade e integração no ecossistema global de repositórios de dados de investigação.

A publicação de dados (*depósito*) no RDM é um processo mediado, sob a responsabilidade de um *data steward* designado. Este assegura que os *datasets* cumprem os padrões institucionais em matéria de qualidade dos metadados, documentação, licenciamento e conformidade com os princípios FAIR, antes da sua disponibilização.

Publicação de Dados de Investigação e Compromissos Institucionais

Princípios Gerais

A publicação e a gestão a longo prazo dos dados de investigação no INESC TEC orientam-se por princípios de responsabilidade científica e pelo cumprimento dos requisitos aplicáveis, tanto externos como institucionais. Neste contexto:

- a) Os dados de investigação resultantes de projetos de I&D financiados com fundos públicos devem cumprir os requisitos de partilha de dados definidos pelas entidades financiadoras.

- b) Os dados que suportam resultados científicos publicados devem respeitar as políticas de dados definidas pelas revistas ou editoras, quando aplicável.
- c) O Serviço de Apoio à Gestão presta apoio aos investigadores na elaboração de DMPs promovendo o alinhamento atempado com os requisitos das entidades financiadoras e a adoção de práticas adequadas de gestão de dados de investigação. Sempre que exigido por financiadores ou editores, os investigadores devem elaborar e manter um *DMP*, com apoio institucional que assegure a conformidade e consistência.
- d) Sempre que apropriado, os *datasets* produzidos no âmbito das atividades de investigação do INESC TEC devem ser disponibilizados através do repositório institucional de dados (*RDM INESC TEC*), como parte da abordagem institucional à gestão, visibilidade e acessibilidade a longo prazo dos dados de investigação.
- e) O INESC TEC reconhece que os *datasets* podem ser depositados em repositórios externos ou infraestruturas especializadas, sempre que tal seja exigido por financiadores, parceiros ou normas disciplinares, ou pela natureza operacional dos dados, desde que seja assegurada a adequada identificação institucional e contextualização.
- f) Todos os *datasets* publicados devem possuir identificadores persistentes, de forma a garantir a sua rastreabilidade, citação e acessibilidade a longo prazo.
- g) Como princípio geral, os *datasets* devem ser disponibilizados em acesso aberto, exceto quando existam restrições legais, éticas ou contratuais (por exemplo, proteção de dados pessoais, confidencialidade industrial, propriedade intelectual sujeita a patente ou acordos de consórcio vinculativos).
- h) O [Regulamento de Propriedade Intelectual do INESC TEC](#) e quaisquer acordos formais entre o INESC TEC e instituições que cedam recursos humanos à instituição prevalecem sobre o disposto na presente política.
- i) O INESC TEC compromete-se a assegurar a preservação digital e a acessibilidade a longo prazo dos *datasets* publicados, incluindo armazenamento seguro, cópias de segurança e manutenção de metadados, garantindo ainda apoio institucional ao longo do processo de publicação de dados.

Orientação para Investigadores

A partilha de dados em acesso aberto, alinhada com os princípios *FAIR*, depende de uma preparação estruturada e de documentação adequada ao longo de todo o ciclo de vida dos dados. Os investigadores devem considerar as necessidades de gestão de dados desde o início dos seus projetos e, sempre que aplicável, elaborar um DMP que oriente este processo. Os DMP desempenham um papel importante na documentação da forma como os dados de investigação

serão geridos ao longo do seu ciclo de vida. Dependendo dos requisitos do projeto, podem abranger aspetos como a organização dos dados, o armazenamento, as condições de acesso, a seleção de repositórios, a publicação, a preservação e as responsabilidades associadas à gestão dos dados.

Para apoiar a publicação e partilha de dados de investigação, o INESC TEC disponibiliza mecanismos de mediação institucional, validação técnica e serviços de apoio. Neste âmbito, aplicam-se as seguintes orientações:

- O Serviço de Apoio à Gestão é responsável pela mediação do processo de publicação de dados no RDM INESC TEC. Os investigadores devem submeter os pedidos de publicação através dos canais internos apropriados.
- O Serviço de Apoio à Gestão presta igualmente apoio à publicação em repositórios externos, em conformidade com os requisitos de financiadores ou editores. Por exemplo, o INESC TEC mantém uma comunidade oficial no Zenodo e pode apoiar a criação de comunidades específicas para projetos, de forma a facilitar a organização, visibilidade e disseminação dos *datasets* associados a iniciativas de investigação, assegurando simultaneamente a adequada identificação e atribuição institucional.
- Como regra geral, os *datasets* depositados no repositório institucional são disponibilizados sob uma licença aberta, preferencialmente *Creative Commons Attribution (CC BY)* ou *Attribution Share-Alike (CC BY-SA)*. Podem ser aplicadas outras licenças quando exigido por financiadores, editores ou normas disciplinares, desde que as condições de reutilização estejam claramente definidas nos metadados. Em situações de co-propriedade, a definição da licença deve resultar de acordo entre as instituições envolvidas.
- Todos os *datasets* destinados a publicação devem incluir:
 - os ficheiros relevantes;
 - metadados completos;
 - documentação adequada (por exemplo, um ficheiro README);
 - informação sobre o licenciamento.
- Após a publicação, é atribuído ao dataset um identificador persistente (DOI), através do serviço *DataCite*, gerido pelo Serviço de Apoio à Gestão.
- A publicação de *datasets* gerados em colaboração deve respeitar os acordos estabelecidos entre as instituições parceiras e refletir o definido no DMP.
- Sempre que os *datasets* envolvam participantes humanos ou incluam dados pessoais ou sensíveis, os investigadores devem consultar o Encarregado de Proteção de Dados (*Data Protection Officer - DPO*) na fase de planeamento, antes da recolha de dados e novamente antes da publicação, caso ocorram alterações relevantes face ao inicialmente previsto, de forma a garantir o tratamento lícito dos dados, a anonimização adequada e a sua divulgação responsável.

- Em casos excepcionais, nomeadamente quando estejam em causa dados pessoais, acordos de confidencialidade ou questões de segurança, os *datasets* podem ser depositados com acesso restrito. Nestas situações, os metadados e quaisquer ficheiros não sensíveis devem, sempre que possível, permanecer acessíveis. Estas exceções devem ser devidamente justificadas, com base em princípios de proporcionalidade e avaliação de risco.

Papéis e Responsabilidades

A tabela seguinte sintetiza os papéis e responsabilidades dos principais intervenientes na publicação e gestão de dados de investigação no INESC TEC. Estes papéis são complementares e aplicam-se ao longo de todo o ciclo de vida dos dados, no âmbito do processo institucional de mediação, assegurando a conformidade com a presente política e o alinhamento com os princípios FAIR.

Interveniente	Responsabilidade	Envolvimento
Investigadores / Equipa de investigação	Assegurar que os <i>datasets</i> cumprem os requisitos desta política; Disponibilizar dados, metadados e documentação completos e adequados.	Ao longo de todo o ciclo de vida dos dados; Responsáveis pelo planeamento, organização, documentação e cumprimento de requisitos éticos, legais e contratuais.
Investigador Responsável (quando aplicável)	Assegurar a conformidade global do projeto com a presente política; Promover que a equipa de investigação cumpre as suas obrigações.	Em todas as fases do projeto; Acompanha o cumprimento da política, incluindo os requisitos associados ao DMP, especialmente nas fases de planeamento e preparação dos dados.
Data Steward (no âmbito do Serviço de Apoio à Gestão)	Mediar o processo de publicação no RDM INESC TEC; Validar a qualidade dos metadados e conformidade com FAIR; Apoiar a publicação em repositórios externos, quando necessário.	Principalmente nas fases de documentação e publicação; Assegura a validação dos <i>datasets</i> (estrutura, metadados e atribuição de licenças), antes da disponibilização; Presta apoio na elaboração de DMPs e acompanha os investigadores ao longo do ciclo de vida dos dados.
Encarregado de Proteção de Dados - DPO	Avalia <i>datasets</i> que envolvam participantes humanos ou dados pessoais ou sensíveis; Assegurar o cumprimento do RGPD e de	Sempre que estejam em causa dados pessoais ou sensíveis; Deve ser consultado antes da recolha de dados e novamente antes da publicação, caso existam

	outros requisitos legais aplicáveis.	alterações relevantes, garantindo a conformidade legal e a adequação das condições de tratamento e divulgação.
Serviço de Administração de Sistemas	Assegurar o funcionamento, a segurança e a preservação a longo prazo da infraestrutura de dados (<i>RDM INESC TEC</i>), incluindo cópias de segurança; Prestar suporte técnico em articulação com o Serviço de Apoio à Gestão.	Ao longo de todo o ciclo de vida dos dados; Garante a fiabilidade da infraestrutura, o armazenamento seguro e a preservação dos <i>datasets</i> .
INESC TEC	Assegurar o compromisso institucional com práticas de dados abertas e alinhadas com os princípios <i>FAIR</i> .	Papel estratégico e transversal a toda a organização.

Conclusão

A Política de Dados Abertos do INESC TEC reafirma o compromisso da instituição com o reforço da qualidade, abertura e impacto da sua investigação, através de uma gestão responsável e da partilha adequada de dados de investigação. Em alinhamento com padrões europeus e internacionais, e com os princípios FAIR, esta política pretende contribuir para uma cultura consistente e sustentável de gestão de dados. Ao estabelecer orientações claras para a produção, documentação, preservação e partilha de dados, esta política apoia os investigadores na disponibilização de dados que possam ser localizáveis, acessíveis, interoperáveis e reutilizáveis, promovendo assim a transparência, a reprodutibilidade e o valor a longo prazo dos dados de investigação.

A sua implementação assenta em infraestruturas dedicadas, apoio especializado e acompanhamento contínuo. Os investigadores são convidados a participar ativamente neste esforço coletivo, contribuindo para que os dados produzidos no INESC TEC sirvam não apenas a comunidade científica, mas também a sociedade em geral. A presente política será revista pelo menos de três em três anos, ou sempre que tal se justifique em função de alterações nas prioridades institucionais, nos requisitos das entidades financiadoras ou no enquadramento legal aplicável.

Poderá ainda ser realizado acompanhamento periódico, com o objetivo de promover o alinhamento contínuo com esta política e assegurar a coerência entre os DMPs e os *datasets* publicados.

Contatos

Campus da FEUP
Rua Dr. Roberto Frias
4200 - 465 Porto, Portugal

+351 222 094 000
www.inesctec.pt
info@inesctec.pt

