

Os Direitos relacionados aos Dados de Investigação na prática de Text and Data Mining

Os impactos na investigação científica

Suziane Batista Marinho

Doutoranda em Ciência da Informação — Universidade de Coimbra

Maria Manuel Borges

Universidade de Coimbra, CEIS20

A ciência produz cada vez mais dados

Dados de Investigação:

“Registos factuais usados como fontes primárias na investigação, e que são geralmente aceites pela comunidade científica como necessários para validar os resultados de investigação.”

— Definição da OECD (2006)

DADOS a são matéria-prima para validar resultados, reproduzir estudos e gerar novas descobertas.

O seu compartilhamento e reutilização oferecem benefícios para a prática científica.



Reprodutibilidade

Permite verificar e repetir resultados.



Transparência

Torna os processos científicos mais verificáveis.



Reutilização

Reduz a duplicação e amplia o valor científico.

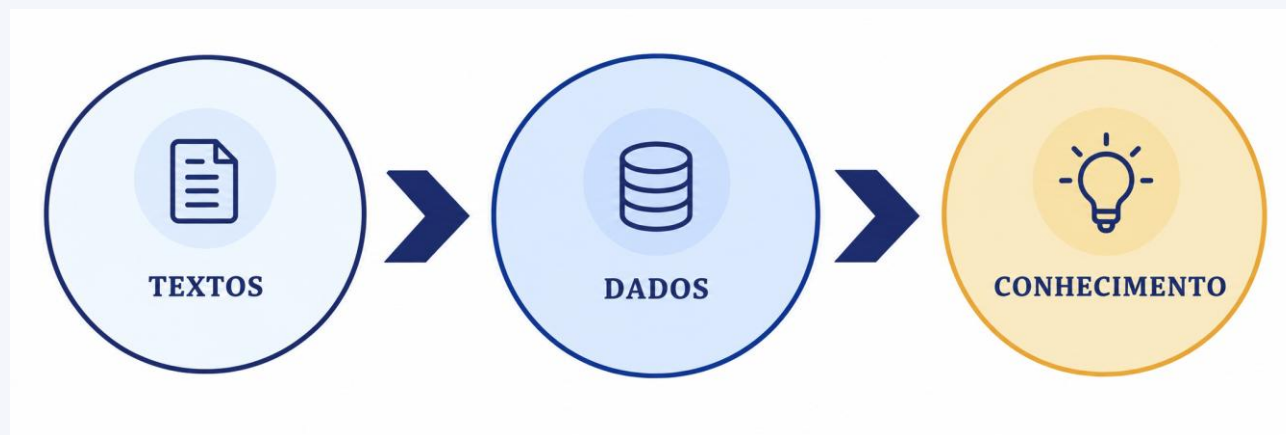
A ciência produz cada vez mais dados

O volume crescente de dados gerados pela investigação científica reforça a necessidade de práticas que permitam a sua exploração eficiente — e o Text and Data Mining (TDM) destaca-se nesse cenário.

TDM

“Qualquer técnica de análise automática destinada à análise de textos e dados em formato digital, a fim de produzir informações, tais como padrões, tendências e correlações.”

Diretiva (UE) 2019/790, art. 2º, nº 2



Proteção jurídica ↔ Ciência aberta

Como os direitos de autor e o direito sui generis das bases de dados influenciam a prática do TDM — e com que impacto na investigação?



Os dados na Investigação

Os dados, por si só, não são protegidos por Direitos de autor. Mas a sua integração em bases de dados ou obras científicas pode introduzir limitações relevantes à reutilização.

IN

Entrada (input)

Os “datasets” usados na investigação.

Se as fontes tiverem direitos de autor, é preciso verificar permissões legais de reutilização — ou obter autorização, com custos e orçamento adicional para licenciamento.

Pergunta jurídica: há acesso e permissão para reutilizar?

OUT

Saída (output)

Os dados gerados como resultado da investigação.

Podem atrair, para si próprios, direitos de autor, conexos ou proteção sui generis — o que também gera questões sobre partilha e reutilização.

Compartilhar exige perceber permissões e condições.

Dados livres, mas nem sempre de acesso livre

Dado isolado

Facto, ato ou informação — não é, por si só, protegido por direitos de autor.



Dado organizado

Ao ser selecionado, estruturado ou apresentado numa obra, pode adquirir proteção.



Dado em base de dados

Pode atrair dupla proteção — Direito de autor e Direito Sui generis.

A proteção das bases de dados científicas

“As bases de dados agilizam o compartilhamento de informações científicas e são fundamentais no processo de investigação”.

(Chapman, 1998)

As bases de dados científicas podem conter quase todos os tipos de dados: artigos, resumos, metadados, dados brutos de investigação. Quando recolhidos numa base não original, o fabricante pode reivindicar o Direito Sui Generis de Banco de Dados.

Diretiva 96/9/CE — dupla camada de proteção

Direito de autor

Protege a estrutura da base: a seleção ou disposição deve representar criação intelectual, originalidade e criatividade do autor.

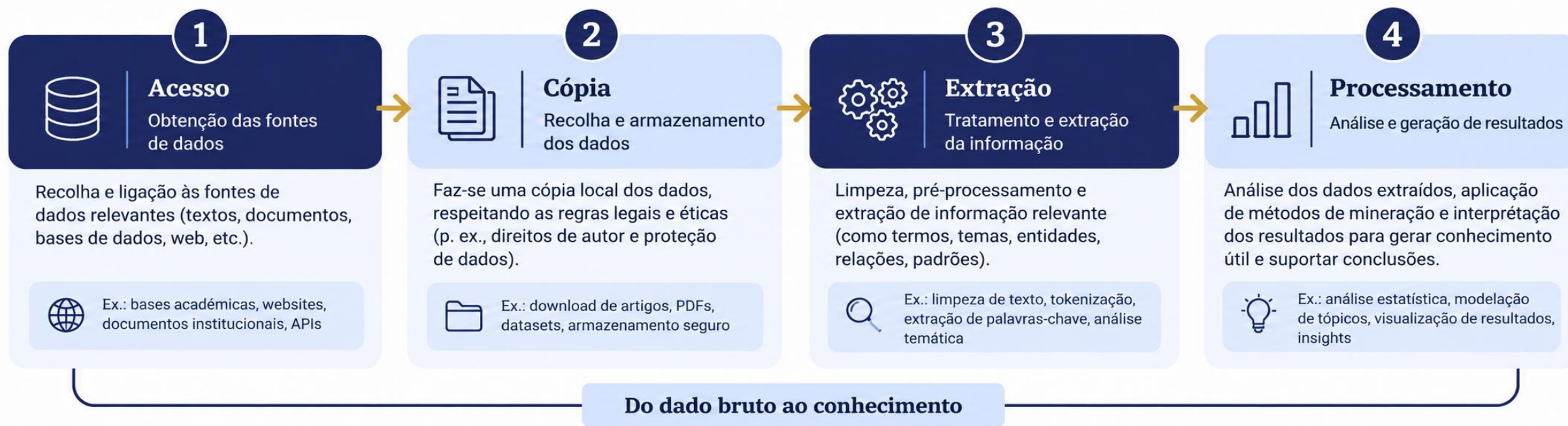
Direito sui generis

Protege quem realizou um investimento substancial na elaboração da base — cumulativo com o direito de autor.

A dupla proteção pode restringir a reutilização, mesmo quando os dados não são protegidos individualmente.

Como o TDM interfere com direitos exclusivos

Para funcionar, o TDM implica várias operações que podem colidir com direitos de titulares de propriedade intelectual:



Na prática: investigadores precisam considerar o acesso aos dados e as condições jurídicas — licenças, restrições contratuais e custos adicionais podem limitar o alcance do TDM.

Do vazio legal à exceção explícita



2001

Diretiva 2001/29/CE

Não havia clareza se o TDM estaria coberto por exceções e limitações — aplicação incerta (Geiger et al., 2018).



2019

Diretiva (UE) 2019/790

Introduz uma exceção explícita e obrigatória para TDM realizado por organismos de investigação, para fins científicos.

Artigo 3.º — a exceção obrigatória para TDM

Os Estados-Membros preveem uma exceção aos direitos no que se refere às **reproduções e extrações** efetuadas por organismos de investigação para a realização de **prospecção de textos e dados de obras** ou outro material protegido **para efeitos de investigação científica**.

Diretiva (UE) 2019/790, art. 3º, nº 1

Um avanço — mas não uma solução completa



O que a exceção favorece

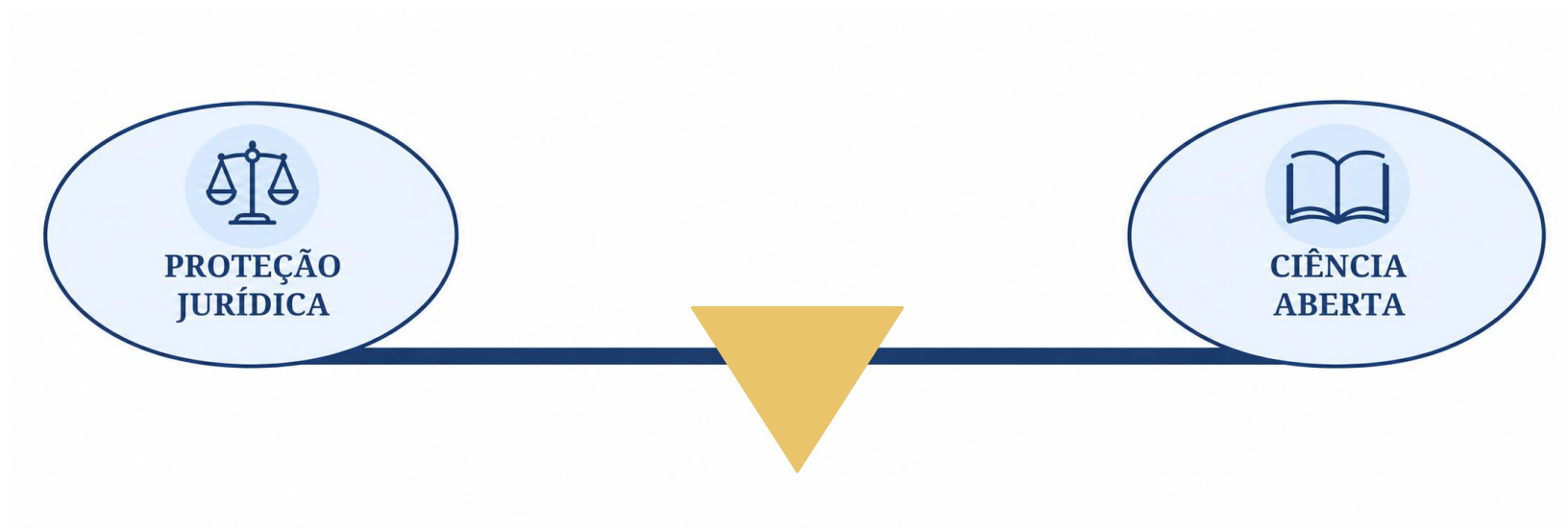
- Harmonização jurídica entre Estados-membros da UE
- Livre acesso aos dados de investigação
- Avanço na construção do conhecimento científico



O que ainda persiste

- Direitos de autor e sui generis ainda podem gerar impacto negativo
- Falta de licenciamento adequado impede a reutilização
- Obstáculos práticos persistem apesar da exceção legal

O ponto de tensão



O desafio é permitir a exploração científica sem ignorar os direitos existentes.

Licenciamento aberto como instrumento de equilíbrio

LICENCIAR PARA REUTILIZAR

Reduz barreiras de acesso, esclarece permissões e favorece a circulação de dados para investigação.

ACESSO



REUTILIZAÇÃO



CONHECIMENTO



AVANÇO CIENTÍFICO

Três ideias para reter

1



Dados, isoladamente, não são em regra protegidos por direitos de autor — mas o enquadramento em obras e bases pode condicionar o uso.

2



O TDM depende de operações de cópia, extração e processamento, tornando relevantes as permissões jurídicas.

3



A Diretiva 2019/790 avança com uma exceção científica, mas licenciamento e acesso aberto continuam decisivos.

PROTEGER E ABRIR NÃO PRECISAM SER OPOSTOS

O desafio está em construir condições jurídicas que permitam explorar o potencial dos dados de investigação e do TDM, contribuindo para uma ciência mais aberta, transparente e acessível.

Referências principais

Chapman, A. R. (1998). A human rights perspective on intellectual property, scientific progress, and access to the benefits of science. WIPO.

Diretiva 96/9/CE do Parlamento Europeu e do Conselho, de 11 de março de 1996, relativa à proteção jurídica das bases de dados.

Diretiva (UE) 2019/790 do Parlamento Europeu e do Conselho, de 17 de abril de 2019, relativa aos direitos de autor e direitos conexos no mercado único digital.

Geiger, C., Frosio, G., & Bulayenko, O. (2018). The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the DSM. SSRN.

Guibault, L., & Wiebe, A. (Eds.). (2013). Safe to be open. Universitätsverlag Göttingen.

Labastida, I., & Margoni, T. (2020). Licensing FAIR Data for Reuse. *Data Intelligence*, 2(1–2), 199–207.

Parolini, G. (2026). Intellectual Property Rights, open science and open questions on research data. *Research Ideas and Outcomes*, 12.

Senftblen, M. (2022). Study on EU copyright and related rights and access to and reuse of data. Publications Office.

Contactos

Suziane Batista Marinho • sbmarinho@student.fl.uc.pt

Maria Manuel Borges • mmb@fl.uc.pt

17ª Conferência Lusófona de Ciência Aberta | CONFOA 2026